Technological Responses to Distracting Media Multitasking in Academic Learning

Dissertation

zur Erlangung des Doktorgrades

der Naturwissenschaften

vorgelegt beim Fachbereich 12, Informatik und Mathematik

der Johann Wolfgang Goethe-Universität

in Frankfurt am Main

von

Daniel Biedermann

aus Frankfurt am Main

Frankfurt, 2023

Vom Fachbereich 12 der
Johann Wolfgang Goethe-Universität als Dissertation angenommen.

Dekan: Prof. Dr. Martin Möller

Gutachter: Prof. Dr. Hendrik Drachsler
Prof Dr. Detlef Krömker

Datum der Disputation: 14.12.2023

German Summary (Zusammenfassung)

Im Zuge der Digitalisierung hat sich das akademische Lernen stark gewandelt. Digitale Medien sind zu allgegenwärtigen Werkzeugen in unserem Alltag geworden und bringen sowohl Chancen als auch Herausforderungen mit sich. Eine dieser Herausforderungen besteht darin, dass der nahezu ununterbrochene Zugriff auf Medieninhalte zu vermehrten Ablenkungen durch sogenanntes Medien-Multitasking führen kann (Dontre, 2021; May & Elder, 2018).

Mediennutzung während oder anstelle des Lernens ist jedoch nicht grundsätzlich problematisch (Shin & Kemps, 2020), denn Mediennutzung umfasst sowohl Aktivitäten mit eher geringem Störungspotenzial, wie z.B. das Hören von Hintergrundmusik, als auch stärker ablenkende Aktivitäten, wie z.B. die Nutzung sozialer Medien (Baumgartner et al., 2017). Entscheidend ist, ob die Mediennutzung das Erreichen der individuellen Lernziele beeinträchtigt. Für die Einordnung von Mediennutzung ist es daher zunächst wichtig, die Mechanismen zu verstehen, die diesem Verhalten zugrunde liegen. Dafür wird in dieser Arbeit das Konzept der Selbstkontrolle verwendet (Duckworth et al., 2019; Inzlicht et al., 2021). Forschungen zur Selbstkontrolle thematisieren den Umgang mit sich widersprechenden Handlungsoptionen, wie etwa dem Fortsetzen des eigenen Lernens oder dem Nachgeben gegenüber digitalen Ablenkungen. Angeborene interindividuelle Unterschiede beeinflussen die Selbstkontrolle ebenso wie intraindividuelle Fluktuationen, zum Beispiel Motivation oder Müdigkeit (Randles et al., 2017). In dieser Arbeit wird beleuchtet, wie Softwaretechnologie dazu beitragen kann, die Konflikte für die Selbstkontrolle zu bewältigen.

Die Arbeit behandelt zwei Teilbereiche: Erstens, die Verbesserung des Datentracking, um Mediennutzung und ihre Auswirkungen auf das Lernen besser zu verstehen. Ein umfassendes Verständnis digitaler Ablenkungen erfordert eine sorgfältige Datenerhebung, um irreführende Daten zu Effekten der Mediennutzung zu vermeiden (Korte, 2020; Parry et al., 2022; Jürgens et al., 2020). Eine Herausforderung ist beispielsweise, dass dieselben Geräte sowohl für das Lernen als auch für die Freizeit genutzt werden. Vereinfachungen wie "mehr Mediennutzung ist gut/schlecht" greifen in solchen Fällen zu kurz, da es schwierig ist, zu unterscheiden, wann Medien für das Lernen und wann für die Freizeit genutzt werden. Es ist daher wichtig zu verstehen, wann und zu welchem Zweck Mediennutzung stattfindet, ob es sich um einen Selbstkontrollkonflikt handelt oder um eine zielgerichtete Handlung. Angesichts der

vielfältigen Einflussfaktoren besteht die Herausforderung zunächst darin, eine Datenerhebungsmethode zu entwickeln, die diese Kontextfaktoren berücksichtigen kann.

Zweitens werden Interventionen behandelt, die das Ziel haben, die Selbstkontrolle bei digitalen Ablenkungen effektiv zu unterstützen. Diese Tools sind in verschiedenen App-Stores in zahlreichen Varianten verfügbar und bieten eine breite Palette an Funktionen, die versprechen, digitale Ablenkungen zu reduzieren (Lyngs et al., 2019). Angesichts der dargebotenen Vielfalt an Funktionen bleibt jedoch unklar, ob diese wirklich alle zu einer Verbesserung der Selbstkontrolle führen. Insbesondere, da es einerseits Tools gibt, die für den Nutzer sehr restriktiv sind, während andere Tools nur die Selbstbeobachtung fördern.

Die beiden Themenbereiche werden in fünf Beiträgen behandelt. Der erste Beitrag diskutiert Anforderungen an Trackingsysteme und stellt eine Implementierung vor, die es erlaubt, Mediennutzungsdaten im Kontext von Lernaktivitäten zu erfassen. Die identifizierten Anforderungen an das Tracking-System umfassen a) die Fähigkeit, eine Verbindung zwischen der Medienaktivität und der entsprechenden Lernaktivität herzustellen. Diese Verbindung erlaubt es auch, die Daten so zu visualisieren, dass ersichtlich wird, welche Lernaktivität zum Zeitpunkt einer Mediennutzung durchgeführt wurde. Das ist wichtig, da eine weitere Anforderung b) die Möglichkeit ist, Informationen über die interne Motivation für die Mediennutzung zu sammeln. Zuverlässig können darüber nur die Nutzer selbst Auskunft geben. Zuletzt ist es auch entscheidend, dass c) ein möglichst breites Spektrum an digitalen Geräten in den Tracking-Prozess einbezogen wird. Dies ist wichtig, um Verzerrungen zu vermeiden, die durch das Ausschließen einzelner Betriebssysteme entstehen könnten (Götz et al., 2017; Jürgens et al., 2020) und berücksichtigt auch die Tatsache, dass Lernende oft mehrere Geräte gleichzeitig nutzen.

In der Implementierung, die als Open-Source Projekt zur Verfügung steht, wurden durch die Integration mit einem Online-Lernsystem die Mediennutzungsdaten mit Lernaktivitäten verbunden. Aktivitäten in dem Online-Lernsystem aktivierten das Medientracking, und erlaubten so Rückschlüsse darauf, welche Mediennutzung bei der Bearbeitung von welchen Lerninhalten stattfand. Die internen Motivationen zur Mediennutzung wurden durch ein Annotationsdashboard abgefragt, das zugleich auch die Möglichkeit zur Datenspende bot (Ohme et al., 2021). In einer begleitenden Studie mit dem Tracking-System wurde ein Datensatz generiert und es wurde überprüft, ob sich die Annahmen zur kontextualisierten Datenerfassung bestätigen. Es zeigte sich eine häufige Nutzung mehrerer Medien, wobei eine lernrelevante Mediennutzung eher auf stationären Computern zu beobachten war, während für Ablenkungen vermehrt Mobilgeräte verantwortlich waren. Weiterhin zeigte sich, dass eine Vielzahl von Apps und Webseiten sowohl für lernrelevante Aufgaben wie auch für persönliche Unterhaltung genutzt wurden.

Der zweite Beitrag präsentiert eine Methode, die frühzeitig erkennen soll, ob ein Text nur unvollständig gelesen wurde, wobei ausschließlich Logdaten zum Scrolling innerhalb des Textes verwendet werden. Gerade längere Texte werden von Studierenden oft nicht vollständig gelesen, sie überspringen längere Passagen, unterbrechen das Lesen durch Mediennutzung (Calderwood et al., 2014) oder brechen das Lesen ganz ab (Forrin et al., 2021).

Die Methode transformiert die vertikale Scrollposition in eine Zeitreihendarstellung, um die Anwendung von Zeitreihenklassifikationsalgorithmen zu ermöglichen. Als Klassenlabel für die Daten wurde ein Indikator entwickelt, der anzeigt, ob ein Text vollständig gelesen wurde oder nicht. Diese Methode wurde mit Scrolldaten von 565 Studierenden getestet und erreichte eine Accuracy von 68%. Die Ergebnisse zeigen, dass die beste Erkennung schon kurz nach Beginn des Lesevorgangs erreicht wird, wodurch das Potenzial für Interventionsstrategien unterstrichen wird. In dem Beitrag werden verschiedene Strategien diskutiert, wie die Genauigkeit dieser Methode in zukünftigen Anwendungen verbessert werden könnte. Insbesondere die Verwendung von Klassenlabeln, die eine genauere Unterscheidung der zugrunde liegenden Gründe für den Abbruch ermöglichen, erscheint als vielversprechende Erweiterung.

Die weiteren Beiträge behandeln Untersuchungen zu digitalen Interventionen zur Verbesserung der Selbstkontrolle.

Zunächst werden im dritten Beitrag die Ergebnisse einer systematischen Literaturrecherche zur Wirksamkeit von digitalen Tools zur Unterstützung der Selbstkontrolle (Digital Self-Control Tools, DSCTs) vorgestellt. Der Fokus lag auf der Anwendbarkeit während des Lernens. Es zeigte sich eine große Vielfalt an unterschiedlichen Features, die helfen sollen, digitale Ablenkungen zu reduzieren. Es wurde deutlich, dass DSCTs mit restriktiveren Features tendenziell wirksamer sind als solche, die hauptsächlich auf die Förderung der Selbstbeobachtung abzielen. In dem Beitrag werden verschiedene Bereiche identifiziert, in denen weitere Untersuchungen erforderlich sind. Erstens konnte aus der Studienlage nicht abgeleitet werden, ob allgemeine Veränderungen in der Mediennutzung tatsächlich zu weniger Ablenkungen beim Lernen führten. Die Teilnehmenden der berücksichtigten Studien waren zwar häufig Studierende, es wurde jedoch meist nicht berichtet, ob diese die DSCTs auch während des Lernens verwendeten. Zweitens wurde in den meisten Studien nur die Aktivität auf einem einzelnen Gerät erfasst. Weniger

Mediennutzung auf einem Gerät könnte möglicherweise zu mehr Nutzung auf einem anderen führen. Diese Beobachtungen zeigen erneut die dringende Notwendigkeit der bereits erwähnten Kontextualisierung von Mediennutzungsdaten.

Im vierten Beitrag wird die Untersuchung zur Rolle von DSCTs während des Lernens vertieft. Eine Befragung von 273 Studierenden lieferte Einblicke in die praktische Anwendung von DSCTs. Es stellte sich heraus, dass Tools, die sich ausschließlich auf Selbstbeobachtung stützen, insbesondere bei einer stark gewohnheitsmäßigen Mediennutzung, als ineffektiv wahrgenommen werden. Des Weiteren gaben die Studierenden über die Gründe Auskunft, aufgrund derer sie die DSCT-Nutzung beendeten. Neben individuell verschiedenen Präferenzen zeigt sich eine situativ variierende Bereitschaft, Einschränkungen in Kauf zu nehmen Auch hier zeigte sich erneut die Herausforderung, eine Balance zwischen zu restriktiven und zu wenig restriktiven DSCTs zu finden, insbesondere bei Plattformen, die sowohl Lerninhalte als auch Unterhaltungsangebote bereitstellen.

Der fünfte Beitrag dieser Arbeit widmet sich der Frage, wie der Aufbau von erwünschten Gewohnheiten mit mobilen Apps besser unterstützt werden kann. Dadurch kann Defiziten in der eigenen Selbstkontrolle entgegengewirkt werden (Galla & Duckworth, 2015), was sich auch positiv auf den Umgang mit Ablenkungen durch Medien auswirken kann (Troll et al., 2020). Mobile Apps können den Prozess der Gewohnheitsbildung durch verschiedene digitale Funktionen fördern (Pinder et al., 2018). Der Beitrag beleuchtet Aspekte der App-basierten Gewohnheitsbildung, die in der bisherigen Forschung eher vernachlässigt wurden. Zum einen sind Kinder die Zielgruppe. Bei Kindern ist eine Verbesserung der Selbstkontrolle erfolgsversprechender als bei Erwachsenen (Coyne & Wright, 2014). Zum anderen wird der Aspekt adressiert, wie die Internalisierung eines Handlungsplans (Gollwitzer, 1999) mit einer möglichst tiefen Verarbeitung erfolgen kann. Ziel der Studie war es, herauszufinden, ob verschiedene digitale Internalisierungsaktivitäten in einer App zu Unterschieden beim Erlernen neuer Handlungspläne führen. Dazu wurde die PROMPT-App entwickelt. Diese enthielt drei verschiedene Aktivitäten, die jeweils eine tiefere oder oberflächlichere Verarbeitung der Handlungspläne fördern sollten. Bei den drei verschiedenen Aktivitäten handelte es sich um ein erneutes Lesen des Plans (passive Aktivität), ein Zusammensetzen des Plans aus einzelnen Wörtern (aktive Aktivität) und eine Aktivität, bei der der Plan durch Emojis repräsentiert werden sollte (konstruktive Aktivität).

Über 27 Tage nutzten 106 Kinder im Alter von 9-14 Jahren die PROMPT-App, um sich täglich wechselnde Pläne mit einer der drei verschiedenen Internalisierungsaktivitäten zu merken, wobei die Aktivitäten alle drei Tage rotierten. Später am Tag sollten die Kinder

versuchen, sich an ihre Pläne zu erinnern. Zusätzlich wurden verschiedene Merkmale der Kinder über Fragebögen und computerbasierte Tests erhoben.
Im Vergleich der Erinnerungsleistung zwischen den drei verschiedenen Aktivitäten zeigten sich differenzierte Resultate. Aktivitäten für eine tiefere Verarbeitung waren nur für Kinder wirksam, die mehr Zeit damit verbrachten. Dies legt nahe, dass es Unterschiede darin gab, wie effektiv Kinder die Internalisierungsaktivitäten nutzen konnten. Zugleich bevorzugten die Kinder die interaktive Aktivität klar vor den anderen Aktivitäten und gaben an, diese gerne häufiger nutzen zu wollen. Die Ergebnisse legen nahe, dass eine kindgerechte Planungs-App personalisiert werden muss, um effektiv zu sein.

Zusammengenommen liefern die Ergebnisse der fünf Beiträge vielfältige und wichtige Erkenntnisse für die technologiebasierte Unterstützung bei digitalen Ablenkungen. Für die beiden Fokusbereiche dieser Arbeit, das Datentracking und die Interventionen, zeigt sich insbesondere die hohe Bedeutung der Berücksichtigung des Kontextes. Beim Datentracking muss der Kontext so umfassend wie möglich erfasst werden, und dabei insbesondere die Lernaktivitäten und die internen Motivationen der Nutzer berücksichtigt werden. Für die noch bestehenden Herausforderungen beim Erfassen dieser Daten werden in dieser Arbeit verschiedene Ansätze zur Lösung diskutiert, etwa eine stärkere Nutzung von Sensoren oder die Analyse von Inhalten im Screenspace.

Auch beim Thema der DSCTs wird deutlich, dass dem Kontext von Daten eine entscheidende Rolle zukommt. In der Arbeit werden verschiedene Beispiele diskutiert, wie Tracking zur Verbesserung von DSCTs eingesetzt werden könnte. Als eine Möglichkeit zeigt sich die kontextbezogene und personalisierte Aktivierung von restriktiven DSCT-Funktionen. Insgesamt bietet diese Arbeit grundlegende Erkenntnisse und konkrete Lösungsansätze, um den Einsatz von Medien im Lernen weniger problematisch zu gestalten. Die Arbeit ebnet den Weg für eine verbesserte Gestaltung und Implementierung von datengesteuerten Tracking-Systemen und digitalen Selbstkontrolltools. Damit trägt sie dazu bei, den negativen Einfluss digitaler Ablenkungen auf den Lernprozess zu mindern und das Potenzial von Medientechnologien für den Bildungssektor stärker auszuschöpfen.

Abstract

With the rise of digitalization and ubiquity of media use, both opportunities and challenges emerge for academic learning. One prevalent challenge is media multitasking, which can become distracting and hinder learning success. This thesis investigates two facets of this issue: the enhancement of data tracking, and the exploration of digital interventions that support self-control.

The *first paper* focuses on digital tracking of media use, as a comprehensive understanding of digital distractions requires careful data collection to avoid misinterpretations. The paper presents a tracking system where media use is linked to learning activities. An annotation dashboard enabled the enrichment of the log data with self-reports. The efficacy of this system was evaluated in a 14-day online course taken by 177 students, with results confirming the initial assumptions about media tracking.

The *second paper* tackles the recognition of whether a text was thoroughly read, an issue brought on by the tendency of students to skip lengthy and demanding texts. A method utilizing scroll data and time series classification algorithms is presented and tested, showing promising results for early recognition and intervention.

The *third paper* presents the results of a systematic literature review on the effectiveness of digital self-control tools in academic learning. The paper identifies gaps in existing research and outlines a roadmap for further research on self-control tools.

The *fourth paper* shares findings from a survey of 273 students, exploring the practical use and perceived helpfulness of DSCTs. The study highlights the challenge of balancing between too restrictive and too lenient DSCTs, particularly for platforms offering both learning content and entertainment. The results also show a special role of media use that is highly habitual.

The *fifth paper* of this work investigates facets of app-based habit building. In a study over 27 days, 106 school-aged children used the specially developed PROMPT-app. The children carried out one of three digital activities each day, each of which was supposed to promote a deeper or more superficial processing of plans. Significant differences regarding the processing of plans emerged between the three activities, and the results suggest that a child-friendly planning application needs to be personalized to be effective.

Overall, this work offers a comprehensive insight into the complexity and potentials of dealing with distracting media usage and shows ways for future research and interventions in this fascinating and ever more important field.

Table of Contents

1 General Introduction

The constant availability of digital media has many advantages in academic learning, such as the ability to research any topic instantly or improved communication between students and teachers. However, having this immediate and ubiquitous access to digital media also increases the potential for frequent distractions through media multitasking (Dontre, 2021). Apps and websites are designed to be highly engaging and habit-forming (Eyal, 2014). This highly engaging content increasingly leads learners to have less attention on the content that is relevant to their learning goals, but rather on digital content that promises instant gratification (May & Elder, 2018).

However, as pointed out by Shin et al. (2020), media use during or instead of learning should not be viewed as inherently problematic. Media multitasking is a very broad concept and includes both activities with minimal interference potential like listening to music, as well as actions that are potentially more disruptive like the use of social media (Baumgartner et al., 2017). Therefore, when assessing the impact of media use on learning, the crucial determinant should be whether such use impedes the attainment of learning objectives.

When dealing with media use that is problematic, it is important to understand the mechanisms that underlie this behavior. To understand the processes involved in conflicts between short-term satisfying impulses (such as entertainment media use) and long-term goals (such as learning success), this thesis refers to the concept of self-control (Inzlicht et al., 2021).

Self-control as an individual trait is largely stable throughout adulthood (Cobb-Clark et al., 2022) and hard to improve. However, self-control also has a state-component, with intra-individual fluctuations due to factors such as motivation, anxiety, stress, or fatigue (Hofmann et al., 2012; Milyavskaya et al., 2019). And it is these situational and contextual factors that also impact distracting media use (Calderwood et al., 2014; Mark et al., 2014; Wiradhany et al., 2021). When we consider the erratic demands of learning, with an often-high demand for self-regulation and oscillations between stressful exam preparation and long bouts of inactivity, it is apparent that the individual's states fluctuate. Thus, when we measure media use of a learner, we cannot fall back to simple statements like "more media use is bad." Instead, it is necessary to consider media use in the context of the learning demands and the internal states of a learner. Thus, the consideration of contextual factors, the differentiation between problematic and less problematic media use, as well as the classification of the underlying factors, will be central to this thesis.

Digital technology does not just create problems. It also offers solutions that can mitigate distractions. As a potential way to address the frequent self-control problems from media multitasking, the field of digital self-control tools (DSCTs) has emerged (Lyngs et al., 2019). These tools, such as website blockers, aim to assist users in managing their self-control struggles. However, the efficacy of these tools remains to be fully substantiated. App stores for various browsers and mobile devices are flooded with tools that claim to reduce digital distractions, yet studies examining their effectiveness rarely focus explicitly on their use during learning. It is crucial to consider how DSCTs should be designed to help students for academic learning, given the complex interactions between the various factors that affect self-control and media multitasking.

This thesis first addresses methodological questions regarding the tracking of behavioral data in a way that can distinguish between different types of media multitasking, and the role of context-awareness. These insights will then inform the investigation of the effectiveness and applicability of digital self-control tools for use in a learning context.

In the following sections, I will first provide the theoretical framing for the five papers that constitute the main body of this thesis. I will then summarize each paper, discuss its main findings, implications, and limitations, and conclude with suggestions for future research directions. This comprehensive overview is intended to provide readers with an understanding of the challenges and potential solutions related to self-control and media multitasking in learning environments.

2 Theoretical Background

Moments of boredom or overwhelm can quickly lead learners to switch to more exciting and entertaining content on their digital devices. Both in controlled environments such as the classroom (Aagaard, 2015) and in self-directed learning settings (Rosen et al., 2013), learners regularly interrupt themselves and others (Pattermann et al., 2022) with digital content, often at intervals of only a few minutes or less. For example, Yeykelis et al. (2014) reported that undergrad students task-switched, on average, every 19 seconds. Similarly, in an observation of learners doing self-paced learning, Calderwood et al. (2014)

observed an average of 35 distractions over the course of three hours of self-directed learning, most frequently to digital media.

Learners have disengaged from their studies since long before the advent of personal digital devices (Vogel-Walcutt et al., 2012). What makes digital distractions more problematic is that developers of digital content use and abuse human psychology to drive user engagement with their content by making them habit-forming (Bayer & LaRose, 2018; Eyal, 2014). Increasing engagement with apps or websites often effectively means increasing the time that users spend on digital content. And, since we as humans can effectively only focus on a single task at a time (Skaugset et al., 2016), the learning content now frequently competes with all the entertaining contemporary digital media offerings that are just one click or one tap away. Since mobile devices are ubiquitous and always available, this form of distraction is not constrained by time or location. Given the importance of effective learning, it is crucial to understand and address the challenges posed by digital distractions, and to be aware of the potential help and remedies that are available.

2.1 Media Multitasking

Media use during learning is studied under the term media multitasking[1], which describes interacting with a digital medium while using another digital medium, or during another non-media activity (May & Elder, 2018; Wang, 2022). It seems that, unlike the capabilities of computers, multitasking here does not mean that a person does the activities truly in parallel. Rather, it is a frequent switching of tasks (Alzahabi & Becker, 2013). The effects of media multitasking during learning have been well-researched in recent years. According to this research, it is frequent among students and appears to impair academic performance. A review of the effect of media multitasking on academic performance by May and Elder (2018) showed interferences of media multitasking on attention, working memory, test performance, reading comprehension, note-taking, and grades. The negative impact of media multitasking specifically for reading tasks has been reviewed by Clinton-Lisell (2021), who found that media multitasking during reading leads to longer reading times and when reading time is limited, also to worse reading comprehension. These detrimental effects occur not only because learners spend less time with the learning

[1] Other frequently used concepts are "internet addiction" (Young, 1998) and "smartphone addiction" (Lanette & Mazmanian, 2018). However, these terms describe a behavioral addiction. This thesis, meanwhile, deals with the everyday, non-pathological type of media multitasking.

content in total, but also because the repeated interruptions prevent sustained attention and deep processing of the content (Liu & Gu, 2020).

Negative emotions such as boredom, fatigue, or anxiety can lead to media multitasking. Some use it as an avoidance strategy against negative emotions (Shin & Kemps, 2020). However, positive emotions such as the enjoyment that is associated with entertainment can also lead to media multitasking (Hwang et al., 2014). This highlights that media multitasking subsumes a variety of different behaviors, with different antecedents. Thus, if we care about reducing the negative impact of digital distractions, it is necessary to specify which type of behavior is problematic.

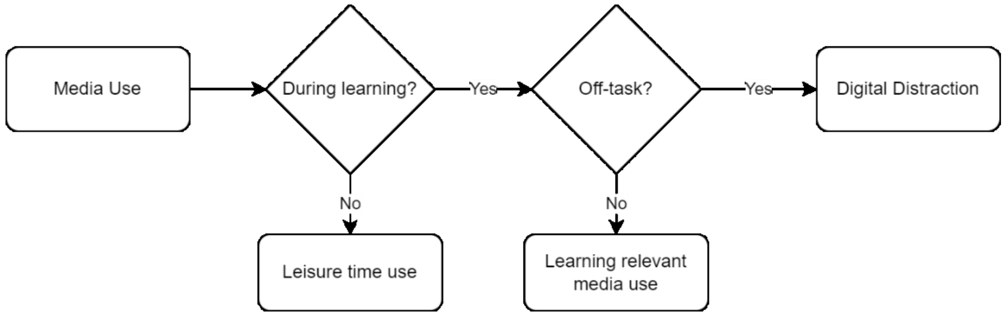
While there are clear situations in which media multitasking during learning has negative effects, there are also many ambiguous types of media use (see figure 1). A paradigmatic example of media use during learning is receiving smartphone notifications (Rozgonjuk et al., 2019). These can be particularly distracting when they appear during learning, but they can also provide valuable information. For example, a notification from a social media app about a new video might be distracting, but a notification from a peer chat could contain the response to a question about the learning materials. One might welcome the latter notification, while the former is a distraction, but one will only know after looking at the notification. Complicating matters is that notifications can become a distraction even if one does not attend to them. Stothart et al. (2015) observed that errors in an attention task increased when participants received mobile notifications that they did not attend to. The researchers hypothesized that the knowledge of having received a notification increased task-irrelevant thoughts and mind wandering, which persists beyond the duration of the notification. In essence, while some instances of media multitasking during learning are clearly detrimental to learning, others are far more ambiguous.

Some forms of media use can even be beneficial during learning. For example some people experience improved learning when listening to music (Li et al., 2021). Nevertheless, listening to music is part of standard measures of media multitasking (Baumgartner et al., 2016), highlighting that measuring media multitasking is not necessarily measuring something that is inherently negative. Using digital media to take deliberate breaks can also be beneficial. In particular, highly conscientious people appear to use the media consciously and purposefully to recover from strenuous work (Mark et al., 2018).

It is evident that the reasons for media multitasking are many and its consequences are not always clear (Xu et al., 2019). Thus, considering that media use is an integral part of learning and that it can also have positive effects on motivation, the aim is not to eliminate media use in general, but to prevent it in those cases where it interferes with goal attainment. The interference of behavior with goal achievement is the domain of self-control, which I will introduce in more detail in the following section.

Figure 1

Different interpretations of media use in academic learning.



2.2 Self-Control Challenges in Learning

Self-control[2] plays a crucial role in situations where individuals face a conflict between a long-term goal and a short-term temptation (Duckworth et al., 2019). For instance, when a student experiences the temptation to check their social media feed while studying, this temptation conflicts with their goal of achieving a good grade in their next exam. This type of conflict is called a self-control conflict. The ability to exert self-control is determined by a combination of trait self-control, which refers to an individual's stable, dispositional level of self-control, and state self-control, which refers to an individual's temporary level of self-control in a particular moment in time. Both components of self-control are important in determining the outcome of a self-control conflict.

Learners with low trait self-control are more likely to use digital media as a distraction and, for example, are more likely to respond immediately to mobile notifications, thereby interrupting themselves more often (Berger et al., 2018). High self-control, on the other

[2] To prevent confusion with regards to terminology, it is useful to mention the concept of self-regulation, which is closely related to self-control but theoretically different from it (Milyavskaya et al., 2019). Self-regulation refers to the ability to determine a desired end state, and to monitor and adjust one's thoughts, emotions, and behaviors to achieve this end state. Self-control involves the ability to control and adapt one's responses to changing circumstances or goals. While theoretically different concepts, they are very closely related. The defining characteristic for self-control is the presence of a conflict between two mutually exclusive actions (Inzlicht et al., 2021).

hand, is correlated with a more intentional type of media use (Reinecke & Hofmann, 2016). As it stands, trait self-control is largely stable in adulthood (Cobb-Clark et al., 2022), and self-control training has only small effects (Friese et al., 2017).

The state component, on the other hand, varies within individuals. State self-control refers to the fluctuations in the capacity to exercise self-control in each situation. In the past, it was believed that self-control was a finite resource on its own, however, recent research suggests that this is not the case (Vohs et al., 2021). Instead, the ability to exert state self-control is influenced by a variety of factors, including cognitive resources such as attention and working memory, emotional arousal, sleep, and fatigue (Inzlicht et al., 2014). The social context can also impact an individual's capacity to exercise self-control in a variety of ways. The presence of others can increase an individual's motivation to control their behavior, as well as increase their perceived accountability (Hofmann et al., 2012), thus increasing self-control in social situations. On the other hand, social pressure can undermine an individual's self-control, particularly when it comes to temptations from media use (Xu et al., 2019). The combination of influence from external and internal factors shows that it is not easy to predict the fluctuations in self-control. Given the complexity of self-control and the influence of several factors, it is important to explore different approaches to exerting self-control.

Self-control has long been viewed primarily as the application of willpower, i.e., just resisting unwanted impulses. However, newer models of self-control (Inzlicht et al., 2021) also see the use of strategies and the proactive preparation for a situation as acting in a self-controlled manner. For example, in the process model of Duckworth et al. (2016), self-control is seen as a dynamic process with multiple stages, including situation modification, attentional deployment, situation appraisal, and response. In each stage, an individual can perform actions to help in the self-control conflict. For example, in the stage of situation modification, individuals can remove the temptations, e.g., by turning off their phones, before starting a study session. In the stage of attentional deployment, individuals can direct their focus either on the task at hand or check their phone regularly. In appraising, they decide to respond to something that they saw on their phone, or to put the phone away again. Finally, in the response stage, individuals either give in to the temptation or resist it, which would be the classical application of willpower. In this model, it ought to be easier for the individual to exercise self-control in earlier phases, for example, not to take the smartphone to the place of learning in the first place (Duckworth et al., 2016). An experience sampling study that prompted the strategies that people used to handle self-control conflicts, suggested that different strategies have different levels of success, and combined strategies were particularly successful (Milyavskaya et al., 2020). A particular promising strategy to reduce the need for willpower is building good habits.

2.3 Habits

Habits are a type of learned behavior that are repeated frequently and tend to occur automatically in response to specific contextual cues (Wood & Rünger, 2016). They play a key role in shaping our behavior and influencing our decision-making processes. Habits are formed through the repetition of behaviors in response to specific contextual cues, and over time, the behavior becomes more automatic, requiring less conscious effort (Galla & Duckworth, 2015). Habits are responsible for a large portion of our everyday behavior (Wood et al., 2002), and they also account for a significant amount of media use (Larose, 2017). If a student has developed a habit of opening a social media site whenever they use their web browser, then this habitual behavior might lead them to do just that, even if their intention were to use their computer for learning. These habit slips can happen due to similar factors as those mentioned previously in relation to media multitasking and self-control: stress, motivation, overwhelm, or fatigue can lead to someone reverting to their habitual behavior instead of their intended behavior (Wood et al., 2014). Although the degree to which media use is habitual varies, it is widespread and especially pervasive with smartphones (Bayer & LaRose, 2018). Oftentimes, users perceive media use that is habitual as particularly meaningless, but at the same time, also particularly difficult to get rid of (Lukoff et al., 2018). However, habits are not set in stone.

While habits themselves are unconscious and automatic, the development of new habits or the modification of existing ones can be controlled to a certain degree (Mazar & Wood, 2018). By building habits that support desired behaviors, individuals can rely on automatic processes to guide their actions, potentially reducing the demands on their self-control (Galla & Duckworth, 2015). For example, students can build a habit of turning off their notifications during learning sessions, and thereby reduce the amount of mental effort required to resist distractions (Troll et al., 2020).

The previous sections have established the problem that we are facing. Media multitasking during learning becomes particularly problematic when it interferes with learning goals and becomes a self-control failure. In the next section, I introduce the topic of digital solutions for digital self-control challenges to explore if, and how, these can be a remedy for the problem.

2.4 The Role of Technology

2.4.1 Tracking Digital Self-Control

As the preceding chapters established, self-control failures can occur due to various contextual factors. These include the student's level of engagement in the learning task, their motivation, and whether the interruption was intentional or a result of habit. A thorough understanding of these factors is critical to address self-control conflicts effectively.

However, these diverse factors pose significant challenges to conventional methods of measuring media use. The traditional method of using self-reports of media use is prone to errors such as recall bias (Schwarz & Oyserman, 2001), and self-reported media use has thus proven to be unreliable and only weakly correlated with actual use (Parry et al., 2021). Hence, there is a clear necessity to use digital tracking methods for measuring media use. Only by utilizing digital tools to monitor media consumption can we obtain a clearer picture of how learners engage with digital media.

But even with digital tracking methods, pitfalls remain. Even when they reduce the problem of recall bias, other biases remain (Jürgens et al., 2020). For example, different populations use different types of digital devices that might not be included in a tracking environment. Furthermore, digital tracking does not automatically entail a regard for the context.

The most important context for the purpose of this thesis is the learning activity. Tracking needs to distinguish between media use during learning and media use during leisure. For example, tracking media use throughout the day might detect a lot of use that is not at all related to distractions from learning; a student may exhibit extensive media use during their leisure time, which warrants different consideration than if the media use occurred predominantly while the student was learning. Understanding these specific circumstances surrounding media use is essential to identify the areas in which learners may struggle with self-control. In light of the increasingly connected and multi-sensory nature of computing, context-awareness (Schilit et al., 1994) presents an opportunity to collect data in a more holistic manner.

Considering the diverse ways in which learners engage with digital media, it is crucial to consider not only the context but also the variety of devices used during the learning process. For example, tracking only data on Android smartphones would disregard the entire population of those who use a smartphone with Apple's iOS operating system. Tracking only one platform not only reduces the sample size, but the sample could also become less representative. For example, small differences in personality between Android and iOS users seem to exist (Götz et al., 2017).

Students also rely on both laptops and smartphones for their educational needs (Carter et al., 2017; Patterson & Patterson, 2017), making it necessary to incorporate multi-device tracking for a comprehensive understanding of media use. However, tracking differs across platforms and operating systems, as each has their own possibilities and limitations for data collection. The iOS operating system, for instance, imposes significant restrictions on usage tracking to third parties. Consequently, researchers have sought alternative tracking methods, such as requesting users to "donate" their usage data for analysis (Baumgartner et al., 2022; Ohme et al., 2021). Such an approach, which relies on proactive users, has different considerations attached to it than an approach where the tracking simply runs as a background service on a device, with barely any user input required. Given that most users have multiple devices, any truly holistic tracking system will need to incorporate multiple platforms, while also having to consider how to integrate the measurement of further contextual factors like the learner motivation or the learning environment.

Given the importance of accurately tracking media use in context as a prerequisite for drawing valid conclusions about media multitasking, a significant portion of this thesis will be dedicated to exploring the complexities and challenges associated with tracking.

2.4.2 Digital Self-Control Tools

Technology also plays a role in alleviating digital distractions. Digital countermeasures are needed because it is unrealistic to expect students to completely disconnect from digital technologies. While banning digital devices from classrooms and lecture halls might be an option (Berry & Westfall, 2015), the impact of such policies reaches its limits when students leave the classroom and learn at other places. The increasing adoption of flipped classrooms and distance learning scenarios (Dziuban et al., 2018) suggests that home-based learning is on the rise. Therefore, it is crucial to develop tools and strategies designed to minimize the detrimental effects of digital distractions in all environments, particularly those where restrictive rules and bans are ineffective. Technological interventions can be devised to help learners manage their self-control struggles. In the subsequent section, I will delve into various techniques and strategies that have the objective of empowering learners to overcome their digital self-control challenges.

Digital self-control tools (DSCTs) aim to assist users in managing digital distractions and can play a crucial role in helping individuals maintain focus during learning. These tools come in a variety of forms and offer a range of features to support users in their efforts to exert self-control over their digital behavior (Lyngs et al., 2019). For instance, website blockers allow users to restrict access to specific websites or applications for designated periods, acting as a digital barrier against potential distractions (Mark et al., 2017). On the other hand, usage visualization tools provide an overview of device usage patterns over time. They aid in comprehending the scope of digital distractions and identifying behavioral tendencies contributing to suboptimal self-control (Y.-H. Kim et al., 2016). Another class of DSCTs includes goal-setting tools, which empower users to create precise targets relating to their device use. The tools then serve reminders and alerts, steering users towards their goals. An example is the "Forest" app, which gamifies the goal-setting process. The app encourages sustained focus by flourishing virtual trees for concentrated efforts, which wither if the user becomes distracted (Parry et al., 2020). These are just a few examples of the wide range of DSCTs that are available for users to choose from (Lyngs et al., 2019). Furthermore, it is worth noting that some digital devices come[3] pre-equipped with such tools. Examples include the "Digital Wellbeing" app on Android devices, and Apple's "Focus" app on iOS devices.

Despite the varied features offered by DSCTs to aid users in controlling their digital distractions, there exist specific challenges related to their usage in a learning context. The example of app and website blockers is a good illustration of these complexities: One could assume that any learner confronted with digital distractions could simply block them all, and thus solve the problem. After all, accessing the distractions is no longer possible. However, while some distracting platforms can indeed be completely blocked, others need to be accessible to the learners. For example, YouTube videos are established in several fields as a source of quality learning resources (Aldallal et al., 2019; Jaffar, 2012). At the same time, YouTube is also a big source of distractions (Moghavvemi et al., 2018).

---

[3] In the beginning of 2023. The ever-changing nature of digital ecosystems might make this statement quickly obsolete.

Thus, a user needs the option to dynamically switch the blocking on and off. This raises the question of how to design the circumvention of restrictions in DSCTs. If it is too easy, the tool might become less effective, and users will start to habitually circumvent their restrictions. If it is too hard, users might find their tool too cumbersome to use (J. Kim et al., 2019). It is unclear how well existing DSCTs address these types of challenges in educational contexts.

However, even without the challenge of these dual-purpose platforms, DSCTs have the limitation that they do not seem to lead to longer-term behavior change, even with prolonged use (Y.-H. Kim et al., 2016; Parry et al., 2020). They may help in the moment, but do not provide a long-term solution. For those seeking a more sustainable approach, cultivating better habits might offer a more promising pathway.

2.4.3 Digital Behavior Change Interventions and Habit Formation

Habit-building apps do not exclusively aim to reduce digital distractions, but they can certainly be used to build habits in this area. Like DSCTs, apps that are designed for habit building incorporate various features and techniques, including notifications, reminders, rewards, and personalization (Pinder et al., 2018; Stojanovic et al., 2020).

A habit-building strategy that seems well-suited for uses in apps due to its structured nature is the concept of "implementation intentions" (Gollwitzer & Sheeran, 2006; Stawarz et al., 2015). These are specific plans framed as "if [context], then [action]". An example relevant to digital distractions might be: "If I start learning, then I will mute my phone." Implementation intentions aim to condition individuals to respond to a specific context cue, thus facilitating habit formation. While their use has been widely supported empirically (Gollwitzer & Sheeran, 2006) there remains a knowledge gap regarding their integration in app-based habit building, specifically within the realm of digital distractions. Consequently, further research could illuminate more effective ways to use technology in bolstering positive habit formation and enhancing self-control.

Nonetheless, it is important to keep in mind that merely installing a habit-building app does not ensure success in building habits. While apps provide *support* for habit formation, individuals must still actively engage in the repetition of the target behavior. Consequently, research continues to explore how apps can optimally support digital behavior change (Pinder et al., 2018). For instance, smartphone notifications, which are often used as reminders for habit repetitions, frequently fail due to users ignoring them (Visuri et al., 2019), and these reminders can also become distractions themselves

(Johannes et al., 2018). Thus, there is a need for investigating other potential affordances that digital technology can offer to support habit development.

In summary, media multitasking, while not inherently detrimental, can evolve into a digital distraction that hampers student learning. It can adversely impact attention during both at-home and in-class learning, negatively affecting academic performance and creating challenges in information retention. While there are digital tools available to aid in self-control, the efficacy of these tools in genuinely fostering learning and the conditions necessary for their success remain unclear.

3 Research Aims

The examination of distracting media multitasking unveils two crucial roles for technology. Firstly, technology must provide a data foundation to discern when media use during learning becomes problematic and when it does not. Secondly, digital tools can aim to alleviate the challenges posed by media distractions during learning.

From the perspective of tracking, an array of influential factors needs to be considered. A pressing issue that emerges is the insufficient differentiation. It is here that the role of technology becomes most salient, assisting in facilitating more nuanced tracking mechanisms. Consequently, this leads to the first research aim that this thesis addresses:

Research Aim Tracking: Explore the use of technology to effectively track media use during learning, while distinguishing it from non-learning media use.

In contrast, the second perspective highlights possible interventions. In this realm, it is apparent that a variety of tools and interventions are already at the disposal of users. However, the focus now shifts to gauging the efficacy of these tools in addressing the problem of digital distractions during academic learning. This leads to the second research aim:

Research Aim Digital Self-Control Tools: Explore how digital self-control tools should be designed and implemented in order to help learners to manage digital distractions during the academic learning process.

4 Contributions

4.1 Paper 1: Contextualized Tracking

Biedermann, D., Ciordas-Hertel, G., Winter, M., Mordel, J., and Drachsler, H. (2023). *Contextualized Logging of On-Task and Off-Task Behavior during Learning*. [Accepted and in the copyediting phase in the Journal of Learning Analytics]

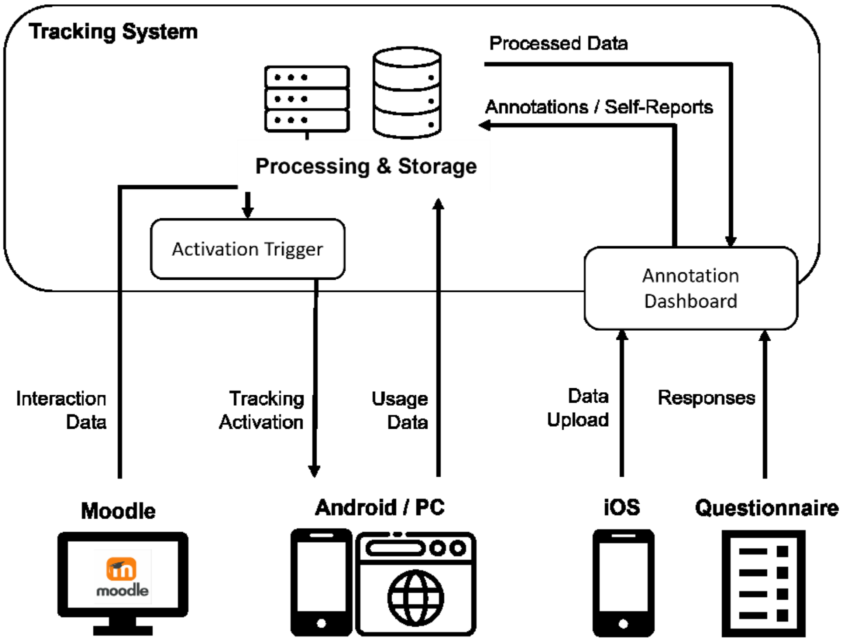
Background: A comprehensive understanding of digital distraction necessitates data tracking that is carried out with diligence to avoid generating misleading data when exploring the impacts of media use (Korte, 2020; Parry et al., 2022). This is especially true when retrospective self-reporting are used, but there are also pitfalls when using digital data tracking methods (Jürgens et al., 2020).

Complications arise when students use the same devices for both leisure and learning, rendering simplified declarations such as "more device use is good/bad" inadequate. Therefore, understanding the context of media use becomes crucial. Given the myriad reasons for media multitasking, the challenge lies in devising a data tracking method that discerns whether media use acts as a distraction or aids learning while also grasping the context surrounding each instance. In response to this challenge, this paper proposes a set of requirements to address this challenge. Based on these guidelines, we developed a tracking system (figure 2) capable of generating highly contextualized data sets. A subsequent study using this system allowed us to evaluate our foundational assumptions on a data set specifically generated for this purpose.

Methods: We employed the tracking system in a study involving 297 students who partook in a fully online Moodle course, which they had to complete within a maximum of 14 days. Throughout the course, we asked students to track their digital activities using the tracking system on their chosen devices. They had the liberty to decide whether and which device to track. Upon completing the learning material at their pace, students could review their data and annotate their motivation behind each instance of media use the tracking system recorded during their learning time. The source code for the tracking system is available at https://gitlab.com/edutex.

Figure 2

High level overview of the components of the tracking system.



Findings: In the reference study, 177 of the 297 participants consented and recorded activity on one or more data sources, with the browser being the most recorded source. Most participants (135) recorded activity on only one data source, while 41 used two and one used three. The system recorded 1387 media activities from all consenting participants, an average of 8.11 activities per participant. Of these, 92% were annotated. Participants' privacy concerns were inferred from the occasions on which they chose 'don't want to answer' as an annotation. This was selected 34 times by 17 different participants, implying a careful consideration of privacy. These two observations suggest that users indeed accepted voluntary annotation. Twenty-four different activities had both on-task and off-task annotations, with varying proportions between categories. Annotations indicating on-task activities were frequently selected for browser activity, while off-task activities were more frequent on smartphones. Examining the time difference between an activity and its annotation, we found that while the average time difference was approximately 13 hours, outliers notably skewed the data. The participants annotated half of the activities within 50 minutes, and 75% within 1 hour and 38 minutes, indicating a rapid annotation process for most activities.

The study's results confirmed our assumptions about the prerequisites for effective tracking and underlined the significance of integrating the learning context and multiple data sources to attain a holistic understanding of digital distractions. The annotations reflected a diversity of devices utilized by learners during their sessions, often with simultaneous use of multiple devices. Additionally, the data revealed several instances of apps and websites serving a dual purpose, functioning as either a learning resource or an off-task distraction.

4.2 Paper 2: Detecting Disengaged Reading

Biedermann, D., Schneider, J., Ciordas-Hertel, G.-P., Eichmann, B., Hahnel, C., Goldhammer, F., & Drachsler, H. (2023). *Detecting the Disengaged Reader—Using Scrolling Data to Predict Disengagement during Reading*. LAK23: 13th International Learning Analytics and Knowledge Conference, 585–591. https://doi.org/10.1145/3576050.3576078

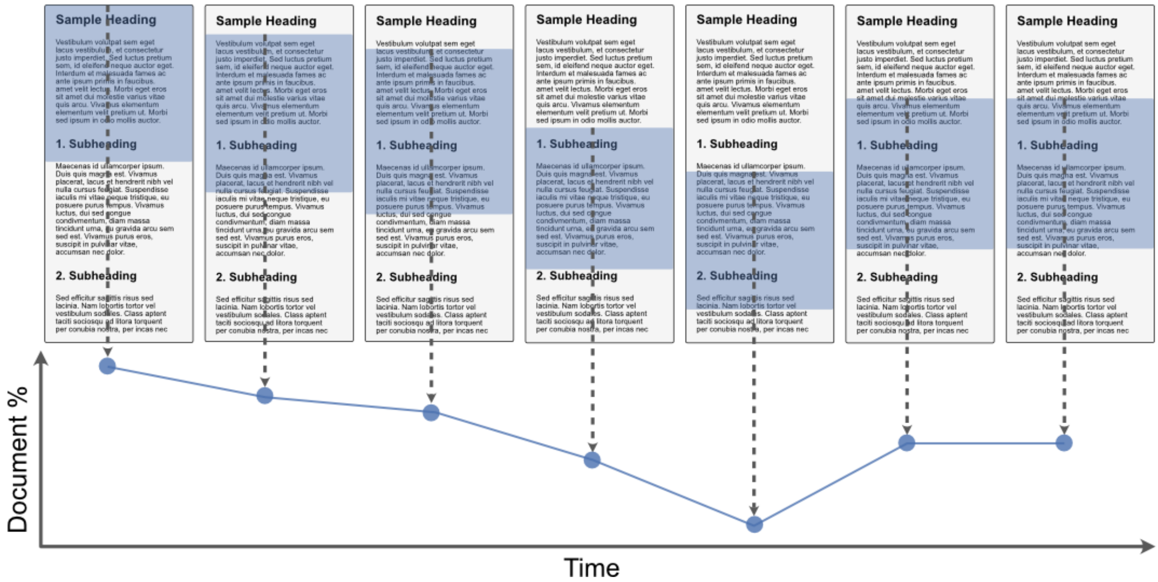
Background: This paper explores the challenge of automatically detecting disengagement during academic reading. Disengagement here means that a student does not read the text in its entirety by skipping passages or abandoning the text before finishing it. Prolonged reading sessions are particularly susceptible to such types of disengagement (Forrin et al., 2021). At the same time, reading long texts is an integral part of higher education curricula. Automatic detection of text-disengagement could help to make students more aware of their patterns and thereby also facilitate the design of interventions to reduce disengagement. In this paper, we present a novel approach to detect disengaged reading using only the scrolling data collected from students in an online course.

Methods: We devised an algorithm that transforms the vertical scrolling position into a time series representation, with a vertical position between 0 (top of the document) and 1 (bottom of the document) at each second (as shown in Figure 3). This transformation of scroll data into a time series format allowed us to perform time series classification on the scroll data. As a classification label, we created an indicator that indicates whether all sections of the text have been read or not. The method was put to the test with the scroll data of 565 students who interacted with eight different texts in an online course. The scrolling data was first transformed into contiguous reading sessions that and then classified.

Furthermore, we conducted an analysis on time series of varying lengths to ascertain the feasibility of making early classifications based on a subset of the scroll data. Additionally, we examined the impact of training on a portion of the texts and then conducting evaluations on the remaining texts, to assess the influence of text selection on the classification performance.

Figure 3

Illustration of the transformation of viewport data into a time series representation. At fixed intervals, the position of the top of the viewport is sampled in relation to the document.



Findings: Our approach yielded an accuracy level of up to 68% in differentiating disengaged reading. We noticed that peak accuracy was achieved approximately 70 seconds into the time series length, beyond which no significant improvement was observed. This finding suggests that early disengagement prediction is feasible with this technique, underscoring its potential usefulness in intervention strategies. The choice of texts for training and evaluation also proved to influence the classification performance; specifically, training on shorter texts led to an increase in false negatives when the method was subsequently applied to longer texts.

The paper discusses various strategies that could potentially enhance the accuracy of this method in future applications. Foremost, it is crucial to identify better labels that can distinguish the different reasons for disengagement.

4.3 Paper 3: Self-Control Tool Literature Review

Biedermann, D., Schneider, J., and Drachsler, H. (2021). *Digital self-control interventions for distracting media multitasking — A systematic review*. Journal of Computer Assisted Learning, 37(5), 1217–1231. https://doi.org/10.1111/jcal.12581

Background: Digital self-control tools (DSCTs) are readily available in abundance across different app stores, employing a wide range of features that promise to mitigate digital distractions (Lyngs et al., 2019). However, given the diversity of different features, it remains unclear whether they all lead to similar levels of improvement. This is especially true given that some features are very restrictive for the user, while other tools use features that only provide small hints as to how a user might improve their own behavior. This paper presents the results of a systematic review of the literature on the effectiveness of digital self-control tools, with a focus on their applicability to academic learning.

Methods: We adhered to the PRISMA guidelines in conducting a systematic literature review, searching for peer-reviewed publications that applied digital interventions aimed at reducing distractions on digital devices, and reported outcome measures related to changes in digital distraction usage. Our analysis encompassed the contents of 16 scientific publications, categorizing them based on the employed features, outcome variables, and participant demographics.

Findings: Our results showed that DSCTs that were more restrictive tended to be more effective, whereas tools that were primarily designed to promote self-awareness often failed to produce significant results. There are also several areas where previous research on DSCTs is still lacking.

First, even though the studies often reported recruiting a student sample, they did not explicitly apply their interventions in a learning context. Thus, it was not clear whether any changes in media use really meant fewer distractions during learning. Second, we found that, with a few exceptions (e.g., Kim et al., 2017; Kovacs et al., 2018), the studies only captured activity on a single device, thereby not doing justice to the multi-device environments that are common in higher education (Gikas & Grant, 2013). A reduction of activity on the smartphone might very well translate to more activity on the notebook, or vice versa. Lastly, most of the research reviewed involved short-term, one-off interventions, leaving the sustainability of DSCT use over longer periods questionable. Considering the existence of novelty effects and potential frustration with restrictions that might arise over time, this diminishes the confidence in the achieved outcomes. These findings serve as a starting point for future research in this field and provide valuable insights into the design and development of digital self-control tools for educational settings.

4.4 Paper 4: Self-Control Tool Survey

Biedermann, D., Kister, S., Breitwieser, J., Weidlich, J., and Drachsler, H. (2023). Use of Digital Self-Control Tools in Higher Education – A Survey Study. [This paper is in its second revision in the journal *Education and Information Technology*]

Background: This paper delves deeper into the use of digital self-control tools (DSCTs) by examining their use during academic learning. Despite the wide range of tools available, there is a lack of understanding about the prevalence and acceptance of DSCTs among higher education students. This survey aimed to explore the features of these tools that are particularly useful for combating habitual media use during learning. This paper provides valuable insights into the use of DSCTs in a higher education context and underscores the need for more research in this area. By exploring the challenges and barriers to habitual use, we hope to inform the design and development of more effective DSCTs that can support students in their learning journeys.

Methods: We conducted a survey of (N = 273) higher education students to gain insights into the use of DSCTs among these students. First, we asked about media use behavior. The participants were asked to answer how often they distract themselves with digital media while learning and how much they suffer from it. The self-report behavioral automaticity index (SRBAI; Gardner et al., 2012) was used to ask about the degree of habituality in media use. For ten different DSCT features, the participants responded about their prior experience with these features, how helpful they perceived them, and if they had previously used a tool with such a feature, why they stopped using it. If participants responded that they had stopped using a DSCT, we used a free text field to ask about the reasons for doing so. Using a qualitative coding approach with three raters, we were able to derive themes that explain why users abandoned their tools. Materials for this study are available at https://osf.io/8zmb7/?view_only=10162da5b5a54e18bd9a7e9830fd0638.

Findings: The perceived helpfulness varied significantly between features. When addressing habitual media use, we found that solely relying on self-tracking proved less helpful for the participants, indicating the need for more comprehensive or restrictive strategies. Another key aspect we delved into were the user testimonials for stopping tool use. These insights illuminated the challenge presented by platforms that serve dual purposes, acting both as a source of distraction and a learning resource. As completely abandoning these platforms isn't feasible for students, they often discontinue DSCTs that impose severe access restrictions. Our survey unearthed that despite the plethora of DSCTs available, many remain relatively unknown, leading most users to resort to pre-installed tools.

4.5 Paper 5: Comparing Internalization Activities for Planning

Biedermann, D.*, Breitwieser, J*., Nobbe, L., Drachsler, H., and Brod, G. (2023). Designing an app to enhance children's planning skills: A case for personalized technology. [This paper is in its second revision in the journal *Behavior and Information Technology*]

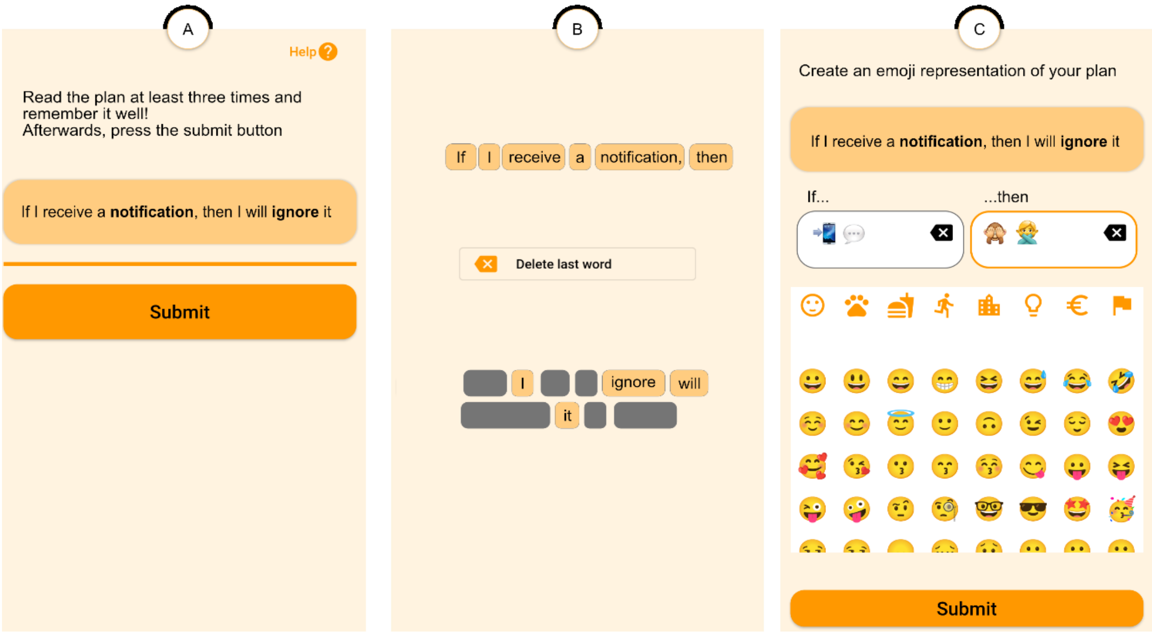
* shared first authorship

Background: This paper focuses on the building of habits via the implementation intention strategy. Conceptualizing a good implementation intention once is not sufficient to attain it. The idea of implementation intentions is that the intended behavior is initiated automatically once the situation is encountered (Webb & Sheeran, 2007), but this necessitates that the mental representation of the plan is sufficiently strong. There is a lack of research comparing these different internalization activities systematically (Hagger et al., 2016), but it is clear that the internalization has to be done with sufficient attention and cognitive resources (Martiny-Huenger et al., 2017). It is certainly not enough to just devise a plan and then never think about it again. Theories of cognitive engagement such as the ICAP framework by Chi et al. (2018) suggest that different types of engagement with an activity should lead to deeper processing, and thus better internalization of a plan, when it is more interactive or constructive rather than just passive. Following this, this study aimed at comparing the effectiveness of different activities of internalizing an implementation intention in a smartphone app.

Methods: We created the PROMPT-app, a planning app to help children create and internalize plans effectively. The app included three different internalization activities that were hypothesized to promote deeper or shallower processing of plans. The three different activities (see figure 4) were a re-reading activity (passive engagement), a puzzle activity (active engagement), and a newly developed emoji activity (constructive engagement). 106 School-aged children (9-14 years) used PROMPT for 27 days in their daily lives. Each morning, they used one of the three internalization activities to internalize one plan. The same internalization activity was used for three consecutive days, before the app switched to the next one. After six hours, the app prompted the children their recall of the daily plan. We used a mixture of computational and manual coding to assess the correctness of the recall. Our pre-registered hypothesis was that the constructive emoji activity would lead to better internalization of plans than using the passive re-reading or the active puzzle activity. Cognitive abilities of the children were assessed with computerized tasks two months after the end of the 27-day period to explore cognitive explanations for inter-individual differences in the effectiveness of the internalization activities. The tasks were an analogical reasoning task and an executive function task that were programmed in "PsychoPy" (Peirce et al., 2019). Scripts and materials for this study are available at https://osf.io/sbcfd/ and the source code for prompt at https://github.com/IndiLab-frankfurt/prompt/.

Figure 4

The three different activities for plan memorization in PROMPT: Re-reading (A), Puzzle (B), Emoji representation (C).



Findings: The findings highlighted the need to support internalization of plans, as children were able to recall plans correctly only about 50% of the time. However, the original hypothesis that constructive internalization is better than active and passive activities did not prove to be true. The constructive processing activity was only more effective for those children who spent more time performing this activity, suggesting that there were differences in how effectively children could make use of the internalization activities. These individual differences were predicted by children's grade level and their analogical reasoning abilities and mediated by time on task.

When asked which activities they would like to use more in the future, the children significantly preferred using only the constructive emoji activity (29.67%), followed by a rotation between different activities (16.48%) and the puzzle activity (15.38%). Considering that motivation to use an app indicates a future intention to use it more (Amaefule et al., 2023), our PROMPT app is also an example of how to make an activity more enjoyable. The findings suggest that a child-appropriate planning app works but needs to be personalized and tailored to children’s learning prerequisites in order to be effective.

5 Discussion

This thesis investigated media multitasking and the associated self-control challenges. The ubiquity of digital distractions in our lives makes this topic particularly pertinent and relevant, and the use of digital tools offers the potential to effectively address and mitigate these challenges. The previous chapters addressed methodological questions, as well as concrete interventions. Both aspects are highly linked, as the methodological considerations are often the limitation to the advancement of better interventions.

5.1 Tracking Digital Distractions: Challenges and Future Directions

To address the first research aim, this thesis explored the possibilities and limits of tracking media use during learning, and Paper 1 shed light on the prerequisites for improving this tracking. It underscored the need to establish a connection between the media activity and the corresponding learning activity, as well as the provision of context information about the motivation for media use. Moreover, it emphasized the importance of including a broad range of digital devices in the tracking process, whenever feasible.

Tracking approaches that do not take these issues into account can lead to simplistic conclusions, especially when observations are simply attributed to 'more media use' or 'less media use'. Given the many factors that influence whether media use in academic learning is harmful or not, these simplistic statements will be misleading. The following subsections will discuss open challenges and future directions about media tracking in detail.

5.1.1 The Challenge of Context for Tracking

As a contribution to the advancement of context-aware media tracking, this thesis presented a system that addressed the issues of balancing self-reporting and behavioral data collection, context-awareness to learning activities, and the inclusion of multiple devices in a single data collection. The evaluation study of the paper demonstrated that such a tracking system can create a highly differentiated dataset. For instance, we discovered a high diversity of devices used by students, and the presence of laptops and smartphones with different operating systems uncovered a wealth of data that would have otherwise not been collected. Moreover, by incorporating means to annotate media use on a per-event basis, the study was able to reveal the frequent occurrence of dual-purpose platforms. These results highlighted that the ability to make these contextual differentiations is essential, and that the interpretation of media activities would otherwise often be inaccurate.

Contextualizing digital activities is the central element for advancing both research and interventions in the field of distracting media multitasking. The limitations of Paper 2 clearly show where contextualized datasets would be appropriate. The label for classifying reading as disengaged was limited to whether a text was fully read or not. With a dataset that contained more information about media use, more nuanced class labels would have been possible. This would allow an investigation into patterns in the scroll data specifically related to media use. The results and in particular the nuanced tracking of Paper 1 was unfortunately not yet available at the time of data collection for Paper 2.

For well-defined and trackable scenarios, Paper 1 has demonstrated how to implement context-awareness, but there are still obstacles that need to be overcome for such a system to work in more general cases. The most pressing problems are the perpetual need to annotate the data and the technical limitations imposed by the different platforms.

5.1.2 The Challenge of Data Annotations

Since self-reporting is always subject to error, the long-term goal is to be able to draw conclusions about the factors that influence media multitasking purely on the basis of behavioral data. However, this has so far only been achieved on an experimental scale. Approaches such as using sensor data to infer emotional states (e.g., Quiroz et al., 2018), are at a stage where they look promising, but not yet good enough to provide truly reliable data (Yadegaridehkordi et al., 2019). Pure behavioral data are not yet sufficient to reliably capture learners' internal states and given that internal states are such a central aspect, they cannot be omitted. Therefore, at this stage, the only way to obtain reliable data is to ask users for it. It is therefore only seemingly contradictory to place so much value on self-reporting while at the same time bemoaning its often lack of validity.

Finding the right timing for the questions always involves trade-offs, and there is no ideal solution. Presenting the annotation question prompts only on specific events (Mehrotra et al., 2015), such as the next phone-unlock (van Berkel et al., 2019), would again increase the time gap between annotation and media use. Moreover, a high frequency and volume of questions could lead to low quality or careless annotations (Eisele et al., 2022). Furthermore, if the annotations immediately follow observed media use, the annotation process itself may introduce an additional distraction. Research into distraction could inadvertently be a source of distraction itself.

There is clearly a trade-off, and the solution from Paper 1 was an annotation dashboard which showed the learners which learning activity they were working on while they used media. This visualization allowed the learners to recall the underlying motivation for media use.

There are concerns that confrontation with one's own media use would lead to reactivity and changed behavior. However, studies with a similar type of approach found very little reactivity (Baumgartner et al., 2022). Furthermore, both Paper 3 and Paper 4 have shown that pure awareness-raising measures are unlikely to lead to quick changes in behavior, and so the risk of causing longer-lasting reactivity effects is low. However, short-term effects cannot be ruled out.

Overall, self-reporting is currently indispensable, even if the medium to long-term goal is to be able to draw inferences from behavioral data alone. Only with reliable annotations can we create datasets and algorithms that allow inferences from behavioral trace data alone.

5.1.3 The Challenge of Platform Restrictions

The annotation dashboard in Paper 1 was also necessary due to the restrictions of iOS operating systems. Only apps developed by Apple are currently able to access information on app usage. Nevertheless, it is indispensable to obtain data from iOS devices as well. In some regions more than half of all mobile devices are iOS devices[4], and excluding these devices would have an impact on the quality of the collected data. We addressed this challenge with a data donation approach (Ohme et al., 2021), where the users could upload their tracking data themselves within the dashboard. While this approach created an additional workload for the users, it also facilitated the collection of fine-grained data on iOS use.

However, even on platforms where app-tracking is possible, the tracked data provides only a limited view of the actions that a user truly performed on their device. For example, with the available tracking APIs on Android operating systems, a system can detect that an app was launched, but not what the app was used for, or how long the app was used for. Thus, if a user launched *App_A* at time *t1*, and *App_B* at *t2*, there is no certainty that the user really used *App_A* for a duration of *t2 - t1*. It might have been open only a fraction of the time. It might have been open, but the user did not use it. As a future way to overcome

[4] https://gs.statcounter.com/os-market-share/mobile/united-states-of-america

this limitation, innovative methods based on screen recordings (Krieter, 2022) have the potential to work around these platform limitations.

5.1.4 Future Direction: Activity Detection on Screen Space

Ideally, we would like to obtain information about what really happened on a user's screen. Screen content analysis might be a future solution for this challenge. In recent years, the capabilities to extract information from screenshots via techniques such as original character recognition and image analysis have reached a level of maturity that enables the automatic classification of the content on screenshots (Chiatti et al., 2017). That is, identifying not only the running applications themselves, but also the actual content that is visible for the user within these applications. A demonstration of these capabilities is the field of "screenomics", which is the analysis of screen recordings over time to create an individual “screenome” out of it (Brinberg et al., 2021; Ram et al., 2020). These screenomes represent the timing and frequency of different types of digital activities. Screenomes provide an overview of an individual's digital activities, i.e., whether they spend their time on social media, or watching videos, or learning. It is apparent how well this relates to the topic of the detection of media multitasking and the associated issues that this thesis discussed.

To obtain this activity data, the device sends screenshots in regular intervals to a backend, where the on-screen activity of the screenshots gets classified by using a combination of machine learning, natural language processing, and human labeling (Brinberg et al., 2021). The alluring advantage of this method is that it focuses on what is truly visible on a screen. It has less abstraction than app-tracking that only reveals part of the whole picture. The disadvantage is the manual annotation task that is still involved (Ram et al., 2020). At least until these systems have sufficiently trained and publicly available models, a high degree of human annotation is necessary.

Human annotation is the bottleneck of this method because annotators have to review thousands of screenshots, even for a few dozen participants (Yee et al., 2022). The sheer requirement for human labor is already a limitation for larger and longer-running research projects. Crucially, however, it also potentially impacts acceptance of such an approach. With API-based app-tracking, one can guarantee to study participants that the analysis of their device activity is limited to aggregated data. The actions of an individual can remain anonymous. This is quite different when there is a manual review of screenshots involved. Chat histories, search entries, etc. would be viewed and annotated by third parties. This process could reveal potentially highly sensitive information, and participants of screenomics studies mention heightened sense of awareness and discomfort as a result (Yee et al., 2022). On the other hand, the recent trajectory regarding the quality of image, object, and text recognition hints at the possibility that the manual sighting of screenshots might only be a transitory necessity. As a perspective for the future of digital activity recognition, the screen analysis approach holds a lot of promise to unify the handling and avoid the complications of limited and ever-changing APIs.

5.1.5 Future Direction: Tracking Learning Outside of Online Environments

A limitation to the data tracking approach of Papers 1 and 2 was that they were restricted to fully online learning environments. While this approach has the advantage of full control over the learning activities, it provides only a partial perspective of the learning process. Students also learn outside of online environments where they also experience digital distractions from media multitasking (I. Kim et al., 2019; Tossell et al., 2015). If a student is writing an essay in a word processor, the browser is only a few clicks away, and a smartphone can be taken to practically every learning environment.

Therefore, just as the tracking of media use needs to be extended to more platforms, the tracking of the learning activities also needs to go beyond what can be easily observed in online learning environments (Blikstein, 2013). This poses new challenges, as tracking activities outside of online environments requires different tracking approaches.

Consider the tracking of students reading text, where the capabilities to track learning activities can vary significantly depending on the medium. As shown in Paper 2, it is possible to collect highly detailed process data in an online learning environment, with information on the exact text length and the scrolling processes and link the trace data to media use. If, however, the same text was accessed as a, e.g., PDF document, the available data would be much more limited. An application tracking approach could detect the point in time when the PDF was opened, and when it was closed, but there would be little information about the processes that occurred in between. Hence, the possible connections between process data and media use would be very scarce. To address this limitation, one could consider making use of the screen content analysis that I discussed in the previous sections. However, this approach also requires that the learning content is viewed on a digital device.

For learning scenarios where the learning content is not on a digital device, sensors could help to detect the learning activities and thus provide context for the media use (Di

Mitri et al., 2018). In this regard, there is potential in the use of smartwatches. Because they are cheap and widespread in use, they can reach a large audience. That is also true for cameras, but they are stationary and thus once again restrict the trackable learning scenarios. Smartwatches, in turn, can be worn in all but the most exotic of learning scenarios. And in learning scenarios where wearing a smartwatch is impossible, media multitasking is also unlikely to be an issue.

With smartwatches, there are promising results where machine learning models succeeded in detecting a variety of day-to-day activities from analysis of the motion and sound sensors. The models can detect activities like writing, typing, or picking up the smartphone (Bhattacharya et al., 2022; Laput & Harrison, 2019). Smartwatch sensors can also yield information about the physical properties of learning environments, like the auditory noise (Ciordas-Hertel et al., 2021). These results indicate that one could use smartwatch sensors to detect whether someone is learning, and therefore also when media use is an interruption of the learning activity.

Sensor data might also be a significant addition to a tracking setup for the detection of internal states. The analysis of smartwatch sensors has shown promise for the detection and identification of emotions (Quiroz et al., 2018). Another avenue could be the use of heart rate variability (HRV), which is the variation in time between successive heart beats. The HRV can be measured reliably using commodity smartwatches (Morresi et al., 2020). There is evidence that HRV is correlated with self-control, with higher HRV being associated with greater self-control (Zahn et al., 2016). These examples show how sensor information could improve data tracking and further reduce the need for self-report data.

5.2 Digital Self-Control Tools: Challenges and Future Directions

The second research aim of this thesis was to explore the ways in which digital tools can alleviate the problems of media multitasking in academic learning. The survey of existing self-control tool research and use was the topic of the Papers 3 and 4. Akin to the considerations on tracking, it became apparent that more consideration of the context is a crucial factor that is missing for DSCTs, both in research and in actual interventions.

Different motivations for media use favor different interventions. This became particularly clear for habitual distractions, for which tools without restrictions are of little help. However, it is not only external factors that favor a particular DSCTs. Different preferences are also a factor that should not be neglected. What one person finds helpful -

another finds annoying. And even within one person, the preference for different interventions can vary.

One fundamental aspect of DSCT use during academic learning that Paper 4 revealed was that a sizable part of students used their DSCTs only during stressful phases, for instance exam preparation periods. Afterwards, they stop using their tool[5]. People are certainly different, and some are critical of their media consumption in general and may thus want to reduce it all the time (Lukoff et al., 2018). For many students, however, the demand for a DSCT is temporary.

Both Paper 3 and Paper 4 also revealed that restrictive tools are not universally applicable in academic learning scenarios. When the distractions occur on dual purpose platforms, access to them cannot be prevented, as that would also entail preventing access to learning content.

Besides DSCTs, there is another way digital tools can support self-control: habit formation. Paper 5 explored the implementation intention strategy in app-based habit building. On the one hand, the study showed that adding more fun input modalities creates a higher motivation to continue using the app. But most importantly, it showed that there are often no one-size-fits-all solutions, and the exciting potential of technology-based interventions is in a high degree of individualization.

5.2.1 The Challenge of Context for DSCTs

The observation from Paper 4 that DSCTs are often used only for stressful phases is consistent with other observations (Lyngs et al., 2022), but it is not consistent with the way DSCTs are often studied. The review from Paper 3 revealed that study outcomes rarely considered these different demands, such as an upcoming exam phase. Thus, it once again becomes apparent that consideration of the context is central, and studies with DSCTs that do not consider the context of media use are at risk of reaching incorrect conclusions about their effects. For instance, consider that a student intentionally takes a break. Media use during this period may not necessarily be a self-control failure and preventing it could lead to more stress for the learner (Mark et al., 2018).

Contextual differences not only affect the interpretation of study results but may also influence the type of DSCT that should be considered appropriate. As shown in Paper 4,

[5] While cramming before an exam may not be the best learning strategy (Dunlosky et al., 2013), helping students to suffer less while cramming is still a worthwhile goal.

there are significant differences between individuals in their preference for self-control tools. For example, students whose media use was strongly habitual found more restrictive tools to be more effective than self-monitoring tools. Although the cross-sectional design of the study did not allow us to examine the influence of internal states such as motivation and fatigue on the effectiveness of DSCTs, it is likely that such states also played a role.

5.2.2 The Challenging Balance of Restrictions

Paper 3 noted that DSCTs tended to be more effective when they used restrictions or sanctions that the users could not easily deactivate or circumvent. This observation was particularly evident in studies that directly manipulated the circumvention difficulty, where increased effort led to greater effectiveness (J. Kim et al., 2019). Conversely, interventions with minimal restrictiveness, such as merely visualizing usage times, often had no impact on reducing distractions (Foulonneau et al., 2016). However, it would be short-sighted to claim that more restrictive interventions are always better. While less restrictive was, on average, less effective, not every user is an average user. Once again, this observation also makes sense from the perspective that media multitasking is not always equally distracting. Sometimes, self-control capacity is high, and a small reminder can be enough, but in other moments, something highly restrictive like a blocker is necessary to stop the distraction.

Highly restrictive interventions also have problems with user acceptance when they limit users' autonomy too much (Lyngs et al., 2022). Reduced acceptance might lead to complete abandonment of a tool. Against this background, the aim should be to use fully restrictive interventions only when they are necessary. Another argument against imposing blanket restrictions is that they are not feasible on some of the widely used dual purpose platforms.

5.2.3 The Challenge of Dual-Purpose Platforms

Dual purpose platforms serve both entertainment and educational purposes and when restrictions prevent access to these platforms, users might perceive restrictions as too excessive. Consider video platforms such as YouTube that can be a significant source of digital distractions and also provide instructional videos (e.g., Aldallal et al., 2019; Jaffar, 2012), or social media and social networking platforms that help as a communication tool for academic discussions (e.g., McPherson et al., 2015; Tartari, 2015). As revealed by the qualitative statements in Paper 4, completely blocking access to these platforms is not feasible for everyone. Learners from this category face a dilemma when attempting to manage their digital distractions while maintaining focus on their studies. On the one hand, they need to access educational content on these platforms. On the other hand, they must also manage the distractions that accompany them.

One suggested solution to this dilemma are the feature modification tools, which are designed to reduce distractions on a platform while still allowing access to its content. They leave core functionality of platforms intact and remove only those parts that a user would consider most distracting. This redesign of platforms has proven successful in studies with changed user interfaces of platforms like Twitter (Zhang et al., 2022), Facebook (Lyngs et al., 2020), or YouTube (Lukoff et al., 2023). In all these studies, addictive and distracting aspects such as news feeds or content recommendations were removed. This gave users more agency, while retaining the option to visit a platform for a specific purpose (Lukoff et al., 2023). Findings from Paper 4 showed that users who have experience with feature removal tools also perceived them as very beneficial. But the results also indicated that these tools are not widely known or used, despite their apparent usefulness. A lack of widespread use may stem from several reasons - these tools might not be adequately advertised, and some users might dislike the removal of parts of the platform. But there is also the fundamental constraint that the commonly used techniques of feature modification are not applicable to mobile devices.

Feature modification tools that are browser extensions overwhelmingly rely on the fact that websites are HTML-based[6], and manipulate the content client-side with JavaScript and CSS rules. Thereby, these tools can easily change or hide parts of a website, and, e.g., remove elements that are particularly distracting. Browser extensions make this process straightforward for websites. For mobile apps, however, these content modifications are far from trivial. Modifying existing mobile apps, which are compiled binaries, is a complex process that involves altering the app's source code and rebuilding it before installation. Even just installing a modified app outside of the app store ecosystem is challenging for casual users, and the process is associated with a host of potential problems, up to bricking the devices (Datta et al., 2022). Consequently, feature modification tools are barely available on smartphones.

[6] There are exceptions like interfaces that are drawn purely in Canvas or in WebGL

5.2.4 Future Direction: Screen Space Feature Modification

Interventions such as “GreaseTerminator”, which applies feature modifications within screen space, present a potential solution to the limitations for feature modification tools (Datta et al., 2022). Akin to the screenomics approach, the GreaseTerminator app streams the smartphone screen to a server backend which classifies the visible content. Beyond only classifying the content, GreaseTerminator also detects areas of the screen which are supposed to be affected by an intervention, for example areas that are particularly distracting and should therefore be hidden for the user. For these areas, the backend sends the information about these areas back to the device which then renders an overlay in the respective areas on top of the original content. Thus, on-screen content can be effectively hidden from the user.

This approach could function across all apps and even cater to the demands of challenging dual-purpose platforms. In addition to the relative platform independence, an exciting aspect of this approach is that it could theoretically react very dynamically to the content on screen. Such an approach could deal with very challenging cases like group chats, where distracting messages (e.g., those containing entertaining images) could be hidden, while information regarding the learning materials should remain visible. Of course, this approach will only work if the screen content classification works reliably. In the previous sections on screen content classification, I have already discussed that there is still some additional effort to be invested in this area.

5.2.5 The Challenges of App-Based Habit Building

Artificial restrictions like completely removing content are not the only way to approach the problem of digital distractions. This is especially true when we consider that distractions most frequently arise from the smartphone (Dontre, 2021) and that the smartphone is just one type of device that students use. In fact, students often prefer to study on their laptops rather than on their smartphones (Learning Innovation, 2022). Thus, while students might not be able to get rid of all their digital devices in their learning environments, they might at least be able to temporarily rid themselves of the most distracting one. For that purpose, beneficial smartphone habits, such as placing the smartphone in different rooms, can be helpful (Troll et al., 2020). Instead of only relying on the crutches of restrictions, support for self-control could also entail building these habits. The benefit is that habits are potentially longer lasting and do not interfere with the autonomy of a user in the same way as a restrictive DSCT. However, this benefit comes at the cost of a process that takes time (Lally et al., 2010). App-based habit building, with its ideas from the area of digital behavior change interventions (DBCI), aims to help motivate and sustain this process (Pinder et al., 2018).

Paper 5 investigated ways to improve the implementation intention strategy for app-based habit building by utilizing lessons from the DBCI literature. In particular, the study highlighted the importance of taking care to make apps not only usable, but also joyful to use. The emoji-activity that we designed specifically for the app was perceived as more enjoyable to use and it was the activity that the participants would prefer to use more in the future. Further analysis of the dataset has also shown that this increased enjoyment results in an increased intention to use of the app (Amaefule et al., 2023). The limitation to the interpretation of the results from Paper 5 against the background of the digital distractions discussed here is that it was not explicitly about media use habits. The paper dealt with the preliminary stage of habit building itself, i.e., planning. However, this does not mean that the results are not applicable to the area of media multitasking. If planned habits are generally easier to put into practice, then this presumably also applies to such habits as smartphone placement. However, this assumption must be validated in future studies.

5.2.6 Future Direction: Context-Aware Triggers for Individualized DSCTs

Another crucial point that emerged in Paper 5 was that not all children benefited equally from the emoji activity. This result highlights that there is no one-size-fits-all approach, and that individualized solutions, which adapt to preferences of the individual, but also to their changing circumstances, are required. In principle, digital applications can be readily personalized. Activating or deactivating functionality via a trigger is an everyday task in user-facing software and has also been tried for DSCTs. The use of different features was[7] at the core of "HabitLab" (Kovacs et al., 2018, 2019). It had a variety of features such as timers, blockers, or feature removal functionality for different websites, and only used a subset of those at a time. In a study which compared a random rotation between different features with a static selection of features, the rotation was more effective, but it also led to more users quitting the tool (Kovacs et al., 2018).

Instead of randomly selecting features, it would therefore make sense for an adaptive DSCT to activate features deliberately. The major difficulty, however, is recognizing when

[7] As of this writing, the project unfortunately appears abandoned.

the trigger should be activated. There is an opportunity here to draw on the findings of the previous chapters on tracking. After all, context awareness could provide the triggers needed to activate features.

Limited forms of context-aware activation are already part of several commercial tools, most prominently in Apple's "Focus" app or the "Digital Wellbeing" app by Google. Both tools can activate restrictions based on the time of day or the location of the user. There are also a handful of studies which investigated context-aware activation, with triggers based on the time of day (Löchtefeld et al., 2013), the location detected in a classroom (I. Kim et al., 2017) or the detection of longer periods without movement (I. Kim et al., 2018). Insights on the actual effectiveness of a context-aware activation, however, remain sparse. Only the "Let's FOCUS" study (I. Kim et al., 2017) compared context-aware activation with manual activation, and found that the participants preferred a location-based activation prompt over a time-based one. Other important questions, such as whether context-awareness leads to more sustained use of DSCTs, and whether it can prevent user attrition, remain, thus far, unanswered.

It is also uncertain whether context-awareness solely based on fixed times or locations can comprehensively accommodate the diverse and complex circumstances inherent to the learning process. An automatic activation during lecture time or when entering a classroom is certainly expedient for the part of learning that occurs within these dedicated locations, but a student studying at home might not want their restrictions activated every time they arrive at home. Neither does every student have a fully regular study schedule that would allow a recurring time-based trigger (Evans et al., 2017; Sher et al., 2022). Being able to study effectively at home is essential, not least with the increase of distance learning and concepts such as flipped classrooms (Akçayır & Akçayır, 2018), and context-aware triggers should be able to handle these realities of academic learning.

A next step for context-aware activation could involve the detection of learning activities from Paper 1. As learning materials become increasingly digital, a student's activity on learning materials could serve as a contextual trigger to activate restrictions. This approach could reduce the need for users to frequently activate and deactivate their tools manually. Combined with the previously proposed solutions for dealing with platform restrictions, these improvements to self-control tools would have the potential to provide significantly better self-control support during learning.

6 Conclusion

This thesis aimed to explore how technology can address digital distractions in academic learning through improvements to media tracking and digital self-control tools. For tracking purposes, a stronger consideration of the contextual factors is necessary to advance insights into media multitasking. The fragmented digital platforms and the multitude of possible influencing factors require dedicated tracking approaches. This thesis presented and discussed an approach for tracking that addresses many of these issues.

For the topic of digital self-control tools, this thesis has shown that it is important to choose the right self-control feature from the many available, taking into account the specific requirements and constraints of the learning situation and the individual learner.

It has been shown that this fit between tools, users and situation is not always right. More individualization and adaptation of self-control features to the users and their situation is needed so that the tools can have their full effect without overburdening or frustrating the users. This observation reveals a connection between both topics of this thesis. The insights gained from tracking of the context have the potential to serve as a basis for improved individualization of self-control tools.

However, for context-aware individualization to succeed, various challenges still need to be solved. Some of these are technical, such as the automatic recognition of internal states, or the fragmented and closed digital ecosystems of operating systems. This thesis highlighted promising research that is advancing these issues, and for some of these challenges, this thesis could directly contribute to the solution.

And these solutions are still urgently needed because digital distractions are unquestionably a challenge that - unless profound legislative or societal changes emerge - is more likely to increase than decrease. The digital products of the attention economy will continue to do everything they can to compete for users' gazes. Fortunately, it also turns out that software has the potential to solve at least some of the problems it creates.

References

Aagaard, J. (2015). Drawn to distraction: A qualitative study of off-task use of educational technology. *Computers & Education*, *87*, 90–97. https://doi.org/10.1016/j.compedu.2015.03.010

Akçayır, G., & Akçayır, M. (2018). The flipped classroom: A review of its advantages and challenges. *Computers & Education*, *126*, 334–345. https://doi.org/10.1016/j.compedu.2018.07.021

Aldallal, S. N., Yates, J. M., & Ajrash, M. (2019). Use of YouTube™ as a self-directed learning resource in oral surgery among undergraduate dental students: A cross-sectional descriptive study. *British Journal of Oral and Maxillofacial Surgery*, *57*(10), 1049–1052. https://doi.org/10.1016/j.bjoms.2019.09.010

Alzahabi, R., & Becker, M. W. (2013). The association between media multitasking, task-switching, and dual-task performance. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(5), 1485–1495. https://doi.org/10.1037/a0031208

Amaefule, C. O., Breitwieser, J., Biedermann, D., Nobbe, L., Drachsler, H., & Brod, G. (2023). Fostering children's acceptance of educational apps: The importance of designing enjoyable learning activities. *British Journal of Educational Technology*, *n/a*(n/a). https://doi.org/10.1111/bjet.13314

Baumgartner, S. E., Lemmens, J. S., Weeda, W. D., & Huizinga, M. (2017). Measuring Media Multitasking: Development of a Short Measure of Media Multitasking for Adolescents. *Journal of Media Psychology*, *29*(2), 1–10. https://doi.org/10.1027/1864-1105/a000167

Baumgartner, S. E., Sumter, S. R., Petkevič, V., & Wiradhany, W. (2022). A Novel iOS Data Donation Approach: Automatic Processing, Compliance, and Reactivity in a

Longitudinal Study. *Social Science Computer Review*, 08944393211071068. https://doi.org/10.1177/08944393211071068

Baumgartner, S., Lemmens, J., Weeda, W., & Huizinga, M. (2016). Measuring Media Multitasking. *Journal of Media Psychology*, *29*, 1–10. https://doi.org/10.1027/1864-1105/a000167

Bayer, J. B., & LaRose, R. (2018). Technology Habits: Progress, Problems, and Prospects. In B. Verplanken (Ed.), *The Psychology of Habit* (pp. 111–130). Springer International Publishing. https://doi.org/10.1007/978-3-319-97529-0_7

Berger, S., Wyss, A. M., & Knoch, D. (2018). Low self-control capacity is associated with immediate responses to smartphone signals. *Computers in Human Behavior*, *86*, 45–51. https://doi.org/10.1016/j.chb.2018.04.031

Berry, M. J., & Westfall, A. (2015). Dial D for Distraction: The Making and Breaking of Cell Phone Policies in the College Classroom. *College Teaching*, *63*(2), 62–71. https://doi.org/10.1080/87567555.2015.1005040

Bhattacharya, S., Adaimi, R., & Thomaz, E. (2022). Leveraging Sound and Wrist Motion to Detect Activities of Daily Living with Commodity Smartwatches. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *6*(2), 42:1-42:28. https://doi.org/10.1145/3534582

Blikstein, P. (2013). Multimodal learning analytics. *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, 102–106. https://doi.org/10.1145/2460296.2460316

Brinberg, M., Ram, N., Yang, X., Cho, M.-J., Sundar, S. S., Robinson, T. N., & Reeves, B. (2021). The idiosyncrasies of everyday digital lives: Using the Human Screenome Project to study user behavior on smartphones. *Computers in Human Behavior*, *114*, 106570. https://doi.org/10.1016/j.chb.2020.106570

Calderwood, C., Ackerman, P. L., & Conklin, E. M. (2014). What else do college students "do" while studying? An investigation of multitasking. *Computers & Education*, *75*, 19–29. https://doi.org/10.1016/j.compedu.2014.02.004

Carter, S. P., Greenberg, K., & Walker, M. S. (2017). The impact of computer usage on academic performance: Evidence from a randomized trial at the United States Military Academy. *Economics of Education Review*, *56*, 118–132. https://doi.org/10.1016/j.econedurev.2016.12.005

Chi, M. T. H., Adams, J., Bogusch, E. B., Bruchok, C., Kang, S., Lancaster, M., Levy, R., Li, N., McEldoon, K. L., Stump, G. S., Wylie, R., Xu, D., & Yaghmourian, D. L. (2018). Translating the ICAP Theory of Cognitive Engagement Into Practice. *Cognitive Science*, *42*(6), 1777–1832. https://doi.org/10.1111/cogs.12626

Chiatti, A., Yang, X., Brinberg, M., Cho, M. J., Gagneja, A., Ram, N., Reeves, B., & Giles, C. L. (2017). Text Extraction from Smartphone Screenshots to Archive in situ Media Behavior. *Proceedings of the Knowledge Capture Conference*, 1–4. https://doi.org/10.1145/3148011.3154468

Ciordas-Hertel, G.-P., Rödling, S., Schneider, J., Di Mitri, D., Weidlich, J., & Drachsler, H. (2021). Mobile Sensing with Smart Wearables of the Physical Context of Distance Learning Students to Consider Its Effects on Learning. *Sensors (Basel, Switzerland)*, *21*. https://doi.org/10.3390/s21196649

Clinton-Lisell, V. (2021). Stop multitasking and just read: Meta-analyses of multitasking's effects on reading performance and reading time. *Journal of Research in Reading*, *44*(4), 787–816. https://doi.org/10.1111/1467-9817.12372

Cobb-Clark, D. A., Kong, N., & Schildberg-Hörisch, H. (2022). The Stability of Self-Control in a Population Representative Study. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4004854

Datta, S., Kollnig, K., & Shadbolt, N. (2022). Mind-proofing Your Phone: Navigating the Digital Minefield with GreaseTerminator. *27th International Conference on Intelligent User Interfaces*, 523–536. https://doi.org/10.1145/3490099.3511152

Di Mitri, D., Schneider, J., Specht, M., & Drachsler, H. (2018). From signals to knowledge: A conceptual model for multimodal learning analytics. *Journal of Computer Assisted Learning*, *34*(4), 338–349. https://doi.org/10.1111/jcal.12288

Dontre, A. J. (2021). The influence of technology on academic distraction: A review. *Human Behavior and Emerging Technologies*, *3*(3), 379–390. https://doi.org/10.1002/hbe2.229

Duckworth, A. L., Gendler, T. S., & Gross, J. J. (2016). Situational Strategies for Self-Control. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *11*(1), 35–55. https://doi.org/10.1177/1745691615623247

Duckworth, A. L., Taxer, J. L., Eskreis-Winkler, L., Galla, B. M., & Gross, J. J. (2019). Self-Control and Academic Achievement. *Annual Review of Psychology*, *70*(1), 373–399. https://doi.org/10.1146/annurev-psych-010418-103230

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving Students' Learning With Effective Learning Techniques: Promising Directions From Cognitive and Educational Psychology. *Psychological Science in the Public Interest*, *14*(1), 4–58. https://doi.org/10.1177/1529100612453266

Dziuban, C., Graham, C. R., Moskal, P. D., Norberg, A., & Sicilia, N. (2018). Blended learning: The new normal and emerging technologies. *International Journal of Educational Technology in Higher Education*, *15*(1), 3. https://doi.org/10.1186/s41239-017-0087-5

Eisele, G., Vachon, H., Lafit, G., Kuppens, P., Houben, M., Myin-Germeys, I., & Viechtbauer, W. (2022). The Effects of Sampling Frequency and Questionnaire

Length on Perceived Burden, Compliance, and Careless Responding in Experience Sampling Data in a Student Population. *Assessment*, *29*(2), 136–151. https://doi.org/10.1177/1073191120957102

Evans, M. D. R., Kelley, P., & Kelley, J. (2017). Identifying the Best Times for Cognitive Functioning Using New Methods: Matching University Times to Undergraduate Chronotypes. *Frontiers in Human Neuroscience*, *11*. https://www.frontiersin.org/articles/10.3389/fnhum.2017.00188

Eyal, N. (2014). *Hooked: How to build habit-forming products*. Portfolio/Penguin.

Forrin, N. D., Mills, C., D'Mello, S. K., Risko, E. F., Smilek, D., & Seli, P. (2021). TL;DR: Longer Sections of Text Increase Rates of Unintentional Mind-Wandering. *The Journal of Experimental Education*, *89*(2), 278–290. https://doi.org/10.1080/00220973.2020.1751578

Foulonneau, A., Calvary, G., & Villain, E. (2016). Stop procrastinating: TILT, time is life time, a persuasive application. *Proceedings of the 28th Australian Conference on Computer-Human Interaction - OzCHI '16*, 508–516. https://doi.org/10.1145/3010915.3010947

Friese, M., Frankenbach, J., Job, V., & Loschelder, D. D. (2017). Does Self-Control Training Improve Self-Control? A Meta-Analysis. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *12*(6), 1077–1099. https://doi.org/10.1177/1745691617697076

Galla, B. M., & Duckworth, A. L. (2015). More than resisting temptation: Beneficial habits mediate the relationship between self-control and positive life outcomes. *Journal of Personality and Social Psychology*, *109*(3), 508–525. https://doi.org/10.1037/pspp0000026

Gardner, B., Abraham, C., Lally, P., & de Bruijn, G.-J. (2012). Towards parsimony in habit measurement: Testing the convergent and predictive validity of an automaticity subscale of the Self-Report Habit Index. *International Journal of Behavioral Nutrition and Physical Activity*, *9*(1), 102. https://doi.org/10.1186/1479-5868-9-102

Gikas, J., & Grant, M. M. (2013). Mobile computing devices in higher education: Student perspectives on learning with cellphones, smartphones & social media. *The Internet and Higher Education*, *19*, 18–26. https://doi.org/10.1016/j.iheduc.2013.06.002

Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *American Psychologist*, *54*, 493–503. https://doi.org/10.1037/0003-066X.54.7.493

Gollwitzer, P. M., & Sheeran, P. (2006). Implementation Intentions and Goal Achievement: A Meta-analysis of Effects and Processes. In *Advances in Experimental Social Psychology* (Vol. 38, pp. 69–119). Academic Press. https://doi.org/10.1016/S0065-2601(06)38002-1

Götz, F. M., Stieger, S., & Reips, U.-D. (2017). Users of the main smartphone operating systems (iOS, Android) differ only little in personality. *PLOS ONE*, *12*(5), e0176921. https://doi.org/10.1371/journal.pone.0176921

Hagger, M. S., Luszczynska, A., de Wit, J., Benyamini, Y., Burkert, S., Chamberland, P.-E., Chater, A., Dombrowski, S. U., van Dongen, A., French, D. P., Gauchet, A., Hankonen, N., Karekla, M., Kinney, A. Y., Kwasnicka, D., Hing Lo, S., López-Roig, S., Meslot, C., Marques, M. M., … Gollwitzer, P. M. (2016). Implementation intention and planning interventions in Health Psychology: Recommendations from the Synergy Expert Group for research and practice. *Psychology & Health*, *31*(7), 814–839. https://doi.org/10.1080/08870446.2016.1146719

Hofmann, W., Baumeister, R. F., Förster, G., & Vohs, K. D. (2012). Everyday temptations: An experience sampling study of desire, conflict, and self-control. *Journal of Personality and Social Psychology*, *102*, 1318–1335. https://doi.org/10.1037/a0026545

Hwang, Y., Kim, H., & Jeong, S.-H. (2014). Why do media users multitask?: Motives for general, medium-specific, and content-specific types of multitasking. *Computers in Human Behavior*, *36*, 542–548. https://doi.org/10.1016/j.chb.2014.04.040

Inzlicht, M., Schmeichel, B. J., & Macrae, C. N. (2014). Why self-control seems (but may not be) limited. *Trends in Cognitive Sciences*, *18*(3), 127–133. https://doi.org/10.1016/j.tics.2013.12.009

Inzlicht, M., Werner, K. M., Briskin, J. L., & Roberts, B. W. (2021). Integrating Models of Self-Regulation. *Annual Review of Psychology*, *72*(1), 319–345. https://doi.org/10.1146/annurev-psych-061020-105721

Jaffar, A. A. (2012). YouTube: An emerging tool in anatomy education. *Anatomical Sciences Education*, *5*(3), 158–164. https://doi.org/10.1002/ase.1268

Johannes, N., Veling, H., Verwijmeren, T., & Buijzen, M. (2018). Hard to Resist?: The Effect of Smartphone Visibility and Notifications on Response Inhibition. *Journal of Media Psychology*, *31*, 1–12. https://doi.org/10.1027/1864-1105/a000248

Jürgens, P., Stark, B., & Magin, M. (2020). Two Half-Truths Make a Whole? On Bias in Self-Reports and Tracking Data. *Social Science Computer Review*, *38*(5), 600–615. https://doi.org/10.1177/0894439319831643

Kim, I., Jung, G., Jung, H., Ko, M., & Lee, U. (2017). Let's FOCUS: Mitigating Mobile Phone Use in College Classrooms. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *1*(3), 1–29. https://doi.org/10.1145/3130928

Kim, I., Kim, R., Kim, H., Kim, D., Han, K., Lee, P. H., Mark, G., & Lee, U. (2019). Understanding smartphone usage in college classrooms: A long-term measurement study. *Computers & Education*, *141*, 103611. https://doi.org/10.1016/j.compedu.2019.103611

Kim, I., Lee, U., & Cha, N. (2018). Exploring Context-aware Proactive Blocking for Distraction Management. *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers - UbiComp '18*, 1202–1207. https://doi.org/10.1145/3267305.3274108

Kim, J., Cho, C., & Lee, U. (2017). Technology Supported Behavior Restriction for Mitigating Self-Interruptions in Multi-device Environments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *1*(3), 64:1-64:21. https://doi.org/10.1145/3130932

Kim, J., Park, J., Lee, H., Ko, M., & Lee, U. (2019). LocknType: Lockout Task Intervention for Discouraging Smartphone App Use. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, 1–12. https://doi.org/10.1145/3290605.3300927

Kim, Y.-H., Jeon, J. H., Choe, E. K., Lee, B., Kim, K., & Seo, J. (2016). TimeAware: Leveraging Framing Effects to Enhance Personal Productivity. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 272–283. https://doi.org/10.1145/2858036.2858428

Korte, M. (2020). The impact of the digital revolution on human brain and behavior: Where do we stand? *Dialogues in Clinical Neuroscience*, *22*(2), 101–111. https://doi.org/10.31887/DCNS.2020.22.2/mkorte

Kovacs, G., Gregory, D. M., Ma, Z., Wu, Z., Emami, G., Ray, J., & Bernstein, M. S. (2019). Conservation of Procrastination: Do Productivity Interventions Save Time Or Just Redistribute It? *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 330:1-330:12. https://doi.org/10.1145/3290605.3300560

Kovacs, G., Wu, Z., & Bernstein, M. S. (2018). Rotating Online Behavior Change Interventions Increases Effectiveness But Also Increases Attrition. *Proceedings of the ACM on Human-Computer Interaction*, *2*(CSCW), 1–25. https://doi.org/10.1145/3274364

Krieter, P. (2022). Are You Still There? An Exploratory Case Study on Estimating Students' LMS Online Time by Combining Log Files and Screen Recordings. *IEEE Transactions on Learning Technologies*, *15*(1), 55–63. https://doi.org/10.1109/TLT.2022.3154828

Lally, P., Jaarsveld, C. H. M. van, Potts, H. W. W., & Wardle, J. (2010). How are habits formed: Modelling habit formation in the real world. *European Journal of Social Psychology*, *40*(6), 998–1009. https://doi.org/10.1002/ejsp.674

Lanette, S., & Mazmanian, M. (2018). The Smartphone "Addiction" Narrative is Compelling, but Largely Unfounded. *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–6. https://doi.org/10.1145/3170427.3188584

Laput, G., & Harrison, C. (2019). Sensing Fine-Grained Hand Activity with Smartwatches. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 338:1-338:13. https://doi.org/10.1145/3290605.3300568

Larose, R. (2017). Media Habits. In *The International Encyclopedia of Media Effects* (pp. 1–9). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118783764.wbieme0190

Learning Innovation. (2022). *Student Device Usage Report April 2022*. The Open University UK. https://www.learninginnovation.info/smart-tech/student-device-usage-report-april-2022-no-animations

Li, F., Wang, Z., Tzi Dong Ng, J., & Hu, X. (2021). Studying with Learners' Own Music: Preliminary Findings on Concentration and Task Load. *LAK21: 11th International Learning Analytics and Knowledge Conference*, 613–619. https://doi.org/10.1145/3448139.3448206

Liu, Y., & Gu, X. (2020). Media multitasking, attention, and comprehension: A deep investigation into fragmented reading. *Educational Technology Research and Development*, *68*(1), 67–87. https://doi.org/10.1007/s11423-019-09667-2

Löchtefeld, M., Böhmer, M., & Ganev, L. (2013). AppDetox: Helping users with mobile app addiction. *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia - MUM '13*, 1–2. https://doi.org/10.1145/2541831.2541870

Lukof, K., Lyngs, U., Shirokova, K., Rao, R., Tian, L., Zade, H., Munson, S., & Hiniker, A. (2023). *SwitchTube: A Proof-of-Concept System Introducing "Adaptable Commitment Interfaces" as a Tool for Digital Wellbeing*. https://doi.org/10.1145/3544548.3580703

Lukoff, K., Lyngs, U., Shirokova, K., Rao, R., Tian, L., Zade, H., Munson, S. A., & Hiniker, A. (2023). SwitchTube: A Proof-of-Concept System Introducing "Adaptable Commitment Interfaces" as a Tool for Digital Wellbeing. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–22. https://doi.org/10.1145/3544548.3580703

Lukoff, K., Yu, C., Kientz, J., & Hiniker, A. (2018). What Makes Smartphone Use Meaningful or Meaningless? *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *2*(1), 22:1-22:26. https://doi.org/10.1145/3191754

Lyngs, U., Lukoff, K., Csuka, L., Slovák, P., Van Kleek, M., & Shadbolt, N. (2022). The Goldilocks level of support: Using user reviews, ratings, and installation numbers to investigate digital self-control tools. *International Journal of Human-Computer Studies*, *166*, 102869. https://doi.org/10.1016/j.ijhcs.2022.102869

Lyngs, U., Lukoff, K., Slovak, P., Binns, R., Slack, A., Inzlicht, M., Van Kleek, M., & Shadbolt, N. (2019). Self-Control in Cyberspace: Applying Dual Systems Theory to a Review of Digital Self-Control Tools. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–18. https://doi.org/10.1145/3290605.3300361

Lyngs, U., Lukoff, K., Slovak, P., Seymour, W., Webb, H., Jirotka, M., Zhao, J., Van Kleek, M., & Shadbolt, N. (2020). “I Just Want to Hack Myself to Not Get Distracted”: Evaluating Design Interventions for Self-Control on Facebook. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–15. https://doi.org/10.1145/3313831.3376672

Mark, G., Czerwinski, M., & Iqbal, S. T. (2018). Effects of Individual Differences in Blocking Workplace Distractions. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI ’18*, 1–12. https://doi.org/10.1145/3173574.3173666

Mark, G., Iqbal, S., & Czerwinski, M. (2017). How blocking distractions affects workplace focus and productivity. *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017*

ACM International Symposium on Wearable Computers, 928–934. https://doi.org/10.1145/3123024.3124558

Mark, G., Wang, Y., & Niiya, M. (2014). Stress and multitasking in everyday college life: An empirical study of online activity. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 41–50. https://doi.org/10.1145/2556288.2557361

Martiny-Huenger, T., Martiny, S., Parks-Stamm, E., Pfeiffer, E., & Gollwitzer, P. (2017). From conscious thought to automatic action: A simulation account of action planning. *Journal of Experimental Psychology: General*. https://digitalcommons.usm.maine.edu/psychology/74

May, K. E., & Elder, A. D. (2018). Efficient, helpful, or distracting? A literature review of media multitasking in relation to academic performance. *International Journal of Educational Technology in Higher Education*, *15*(1), 13. https://doi.org/10.1186/s41239-018-0096-z

Mazar, A., & Wood, W. (2018). Defining Habit in Psychology. In B. Verplanken (Ed.), *The Psychology of Habit* (pp. 13–29). Springer International Publishing. https://doi.org/10.1007/978-3-319-97529-0_2

McPherson, M., Budge, K., & Lemon, N. (2015). New practices in doing academic development: Twitter as an informal learning space. *International Journal for Academic Development*, *20*(2), 126–136. https://doi.org/10.1080/1360144X.2015.1029485

Mehrotra, A., Vermeulen, J., Pejovic, V., & Musolesi, M. (2015). Ask, but don't interrupt: The case for interruptibility-aware mobile experience sampling. *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International*

Symposium on Wearable Computers, 723–732. https://doi.org/10.1145/2800835.2804397

Milyavskaya, M., Berkman, E. T., & De Ridder, D. T. D. (2019). The many faces of self-control: Tacit assumptions and recommendations to deal with them. *Motivation Science*, *5*, 79–85. https://doi.org/10.1037/mot0000108

Milyavskaya, M., Saunders, B., & Inzlicht, M. (2020). Self-control in daily life: Prevalence and effectiveness of diverse self-control strategies. *Journal of Personality*. https://doi.org/10.1111/jopy.12604

Moghavvemi, S., Sulaiman, A., Jaafar, N. I., & Kasem, N. (2018). Social media as a complementary learning tool for teaching and learning: The case of youtube. *The International Journal of Management Education*, *16*(1), 37–42. https://doi.org/10.1016/j.ijme.2017.12.001

Morresi, N., Casaccia, S., Sorcinelli, M., Arnesano, M., & Revel, G. M. (2020). Analysing performances of Heart Rate Variability measurement through a smartwatch. *2020 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 1–6. https://doi.org/10.1109/MeMeA49120.2020.9137211

Ohme, J., Araujo, T., de Vreese, C. H., & Piotrowski, J. T. (2021). Mobile data donations: Assessing self-report accuracy and sample biases with the iOS Screen Time function. *Mobile Media & Communication*, *9*(2), 293–313. https://doi.org/10.1177/2050157920959106

Parry, D. A., Davidson, B. I., Sewall, C. J. R., Fisher, J. T., Mieczkowski, H., & Quintana, D. S. (2021). A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. *Nature Human Behaviour*, *5*(11), 1535–1547. https://doi.org/10.1038/s41562-021-01117-5

Parry, D. A., Fisher, J. T., Mieczkowski, H., Sewall, C. J. R., & Davidson, B. I. (2022). Social media and well-being: A methodological perspective. *Current Opinion in Psychology*, *45*, 101285. https://doi.org/10.1016/j.copsyc.2021.11.005

Parry, D. A., le Roux, D. B., & Bantjes, J. R. (2020). Testing the feasibility of a media multitasking self-regulation intervention for students: Behaviour change, attention, and self-perception. *Computers in Human Behavior*, *104*, 106182. https://doi.org/10.1016/j.chb.2019.106182

Pattermann, J., Pammer, M., Schlögl, S., & Gstrein, L. (2022). Perceptions of Digital Device Use and Accompanying Digital Interruptions in Blended Learning. *Education Sciences*, *12*(3), Article 3. https://doi.org/10.3390/educsci12030215

Patterson, R. W., & Patterson, R. M. (2017). Computers and productivity: Evidence from laptop use in the college classroom. *Economics of Education Review*, *57*, 66–79. https://doi.org/10.1016/j.econedurev.2017.02.004

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pinder, C., Vermeulen, J., Cowan, B. R., & Beale, R. (2018). Digital Behaviour Change Interventions to Break and Form Habits. *ACM Trans. Comput.-Hum. Interact.*, *25*, 15–15. https://doi.org/10.1145/3196830

Quiroz, J. C., Geangu, E., & Yong, M. H. (2018). Emotion Recognition Using Smart Watch Sensor Data: Mixed-Design Study. *JMIR Mental Health*, *5*(3), e10153. https://doi.org/10.2196/10153

Ram, N., Yang, X., Cho, M.-J., Brinberg, M., Muirhead, F., Reeves, B., & Robinson, T. N. (2020). Screenomics: A New Approach for Observing and Studying Individuals'

Digital Lives. *Journal of Adolescent Research*, *35*(1), 16–50. https://doi.org/10.1177/0743558419883362

Randles, D., Harlow, I., & Inzlicht, M. (2017). A pre-registered naturalistic observation of within domain mental fatigue and domain-general depletion of self-control. *PLOS ONE*, *12*(9), e0182980. https://doi.org/10.1371/journal.pone.0182980

Reinecke, L., & Hofmann, W. (2016). Slacking Off or Winding Down? An Experience Sampling Study on the Drivers and Consequences of Media Use for Recovery Versus Procrastination. *Human Communication Research*, *42*, n/a-n/a. https://doi.org/10.1111/hcre.12082

Rosen, L. D., Mark Carrier, L., & Cheever, N. A. (2013). Facebook and texting made me do it: Media-induced task-switching while studying. *Computers in Human Behavior*, *29*(3), 948–958. https://doi.org/10.1016/j.chb.2012.12.001

Rozgonjuk, D., Elhai, J. D., Ryan, T., & Scott, G. G. (2019). Fear of missing out is associated with disrupted activities from receiving smartphone notifications and surface learning in college students. *Computers & Education*, *140*, 103590. https://doi.org/10.1016/j.compedu.2019.05.016

Schilit, B., Adams, N., & Want, R. (1994). Context-Aware Computing Applications. *1994 First Workshop on Mobile Computing Systems and Applications*, 85–90. https://doi.org/10.1109/WMCSA.1994.16

Schwarz, N., & Oyserman, D. (2001). Asking Questions About Behavior: Cognition, Communication, and Questionnaire Construction. *American Journal of Evaluation*, *22*(2), 127–160. https://doi.org/10.1177/109821400102200202

Sher, V., Hatala, M., & Gašević, D. (2022). When Do Learners Study?: An Analysis of the Time-of-Day and Weekday-Weekend Usage Patterns of Learning Management

Systems from Mobile and Computers in Blended Learning. *Journal of Learning Analytics*, *9*(2), 1–23. https://doi.org/10.18608/jla.2022.6697

Shin, M., & Kemps, E. (2020). Media multitasking as an avoidance coping strategy against emotionally negative stimuli. *Anxiety, Stress, & Coping*, *33*(4), 440–451. https://doi.org/10.1080/10615806.2020.1745194

Shin, M., Linke, A., & Kemps, E. (2020). Moderate amounts of media multitasking are associated with optimal task performance and minimal mind wandering. *Computers in Human Behavior*, *111*, 106422. https://doi.org/10.1016/j.chb.2020.106422

Skaugset, L. M., Farrell, S., Carney, M., Wolff, M., Santen, S. A., Perry, M., & Cico, S. J. (2016). Can You Multitask? Evidence and Limitations of Task Switching and Multitasking in Emergency Medicine. *Annals of Emergency Medicine*, *68*(2), 189–195. https://doi.org/10.1016/j.annemergmed.2015.10.003

Stawarz, K., Cox, A. L., & Blandford, A. (2015). Beyond Self-Tracking and Reminders: Designing Smartphone Apps That Support Habit Formation. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, 2653–2662. https://doi.org/10.1145/2702123.2702230

Stojanovic, M., Grund, A., & Fries, S. (2020). App-Based Habit Building Reduces Motivational Impairments During Studying – An Event Sampling Study. *Frontiers in Psychology*, *11*, 167. https://doi.org/10.3389/fpsyg.2020.00167

Stothart, C., Mitchum, A., & Yehnert, C. (2015). The Attentional Cost of Receiving a Cell Phone Notification. *Journal of Experimental Psychology. Human Perception and Performance*. https://doi.org/10.1037/xhp0000100

Tartari, E. (2015). The Use of Social Media for Academic Purposes in Student' Learning Process. *Academic Journal of Interdisciplinary Studies*. https://doi.org/10.5901/ajis.2015.v4n2p393

Tossell, C. C., Kortum, P., Shepard, C., Rahmati, A., & Zhong, L. (2015). You can lead a horse to water but you cannot make him learn: Smartphone use in higher education. *British Journal of Educational Technology*, *46*(4), 713–724. https://doi.org/10.1111/bjet.12176

Troll, E. S., Friese, M., & Loschelder, D. D. (2020). How students' self-control and smartphone-use explain their academic performance. *Computers in Human Behavior*, 106624. https://doi.org/10.1016/j.chb.2020.106624

van Berkel, N., Goncalves, J., Lovén, L., Ferreira, D., Hosio, S., & Kostakos, V. (2019). Effect of experience sampling schedules on response rate and recall accuracy of objective self-reports. *International Journal of Human-Computer Studies*, *125*, 118–128. https://doi.org/10.1016/j.ijhcs.2018.12.002

Visuri, A., van Berkel, N., Okoshi, T., Goncalves, J., & Kostakos, V. (2019). Understanding smartphone notifications' user interactions and content importance. *International Journal of Human-Computer Studies*, *128*, 72–85. https://doi.org/10.1016/j.ijhcs.2019.03.001

Vogel-Walcutt, J. J., Fiorella, L., Carper, T., & Schatz, S. (2012). The Definition, Assessment, and Mitigation of State Boredom Within Educational Settings: A Comprehensive Review. *Educational Psychology Review*, *24*(1), 89–111. https://doi.org/10.1007/s10648-011-9182-7

Vohs, K. D., Schmeichel, B. J., Lohmann, S., Gronau, Q. F., Finley, A. J., Ainsworth, S. E., Alquist, J. L., Baker, M. D., Brizi, A., Bunyi, A., Butschek, G. J., Campbell, C., Capaldi, J., Cau, C., Chambers, H., Chatzisarantis, N. L. D., Christensen, W. J., Clay, S. L., Curtis, J., … Albarracín, D. (2021). A Multisite Preregistered Paradigmatic Test of the Ego-Depletion Effect. *Psychological Science*, *32*(10), 1566–1581. https://doi.org/10.1177/0956797621989733

Wang, C. (2022). Comprehensively Summarizing What Distracts Students from Online Learning: A Literature Review. *Human Behavior and Emerging Technologies*, *2022*, e1483531. https://doi.org/10.1155/2022/1483531

Webb, T. L., & Sheeran, P. (2007). How do implementation intentions promote goal attainment? A test of component processes. *Journal of Experimental Social Psychology*, *43*, 295–302. https://doi.org/10.1016/j.jesp.2006.02.001

Wiradhany, W., Baumgartner, S., & de Bruin, A. (2021). Exploitation–Exploration Model of Media Multitasking. *Journal of Media Psychology*, *33*(4), 169–180. https://doi.org/10.1027/1864-1105/a000303

Wood, W., Labrecque, J., Lin, P.-Y., & Rünger, D. (2014). Habits in dual process models. In *Dual-process theories of the social mind* (pp. 371–385).

Wood, W., Quinn, J. M., & Kashy, D. A. (2002). Habits in everyday life: Thought, emotion, and action. *Journal of Personality and Social Psychology*, *83*, 1281–1297. https://doi.org/10.1037/0022-3514.83.6.1281

Wood, W., & Rünger, D. (2016). Psychology of Habit. *Annual Review of Psychology*, *67*(1), 289–314. https://doi.org/10.1146/annurev-psych-122414-033417

Xu, S., Wang, Z., & Woods, K. (2019). Multitasking and Dual Motivational Systems: A Dynamic Longitudinal Study. *Human Communication Research*, *45*(4), 371–394. https://doi.org/10.1093/hcr/hqz009

Yadegaridehkordi, E., Noor, N. F. B. M., Ayub, M. N. B., Affal, H. B., & Hussin, N. B. (2019). Affective computing in education: A systematic review and future research.

Computers & Education, *142*, 103649. https://doi.org/10.1016/j.compedu.2019.103649

Yee, A. Z. H., Yu, R., Lim, S. S., Lim, K. H., Dinh, T. T. A., Loh, L., Hadianto, A., & Quizon, M. (2022). ScreenLife Capture: An open-source and user-friendly framework for collecting screenomes from Android smartphones. *Behavior Research Methods*. https://doi.org/10.3758/s13428-022-02006-z

Yeykelis, L., Cummings, J. J., & Reeves, B. (2014). Multitasking on a Single Device: Arousal and the Frequency, Anticipation, and Prediction of Switching Between Media Content on a Computer. *Journal of Communication*, *64*(1), 167–192. https://doi.org/10.1111/jcom.12070

Young, K. S. (1998). Internet addiction: The emergence of a new clinical disorder. *CyberPsychology & Behavior*, *1*, 237–244. https://doi.org/10.1089/cpb.1998.1.237

Zahn, D., Adams, J., Krohn, J., Wenzel, M., Mann, C. G., Gomille, L. K., Jacobi-Scherbening, V., & Kubiak, T. (2016). Heart rate variability and self-control—A meta-analysis. *Biological Psychology*, *115*, 9–26. https://doi.org/10.1016/j.biopsycho.2015.12.007

Zhang, M. R., Lukoff, K., Rao, R., Baughan, A., & Hiniker, A. (2022). Monitoring Screen Time or Redesigning It?: Two Approaches to Supporting Intentional Social Media Use. *CHI Conference on Human Factors in Computing Systems*, 1–19. https://doi.org/10.1145/3491102.3517722

Appendix

A: Original Manuscripts

Paper 1

Biedermann, D., Ciordas-Hertel, G., Winter, M., Mordel, J., and Drachsler, H. (2023). *Contextualized Logging of On-Task and Off-Task Behavior during Learning*. Accepted in the Journal of Learning Analytics

Contextualized Logging of On-Task and Off-Task Behaviours During Learning

Daniel Biedermann,[1] George-Petru Ciordas-Hertel,[2] Marc Winter,[3] Julia Mordel,[4] and Hendrik Drachsler[5]

Abstract

Learners use digital media during learning for a variety of reasons. Sometimes media use can be considered “on-task,” e.g., to perform research or to collaborate with peers. In other cases, media use is “off-task,” meaning that learners use content unrelated to their current learning task. Given the well-known problems with self-reported data (incomplete memory, distorted perceptions, subjective attributions), exploring on-task and off-task usage of digital media in learning scenarios requires logging activity on digital devices. However, we argue that logging on- and off-task behaviour has challenges that are rarely addressed. First, logging must be active only during learning. Second, logging represents a potential invasion of privacy. Third, logging must incorporate multiple devices simultaneously to take the reality of media multitasking into account. Fourth, logging alone is insufficient to reveal what prompted learners to switch to a different digital activity. To address these issues, we present a contextually activated logging system that allows users to inspect and annotate the observed activities after a learning session. Data from a formative study show that our system works as intended, and furthermore supports our assumptions about the diverse intentions of media use in learning. We discuss the implications for learning analytics.

Notes for Practice

- Activities on digital devices are an integral part of daily learning, but there is insufficient ecologically valid data on digital device use during learning.
- Existing tracking solutions are often insufficient because they do not allow contextualization of the log data to the learning context and to the multi-device realities of academic learning.
- Our system addresses these issues and allows for the creation of rich data that combines user annotations and logging contextualized to learning activities.

- Keywords: Off-task detection, logging, tools, distractions
- Submitted: dd/mm/yy — Accepted: dd/mm/yy — Published: dd/mm/yy

Corresponding author [1] Email: d.biedermann@dipf.de Address: DIPF | Leibniz-Institute for Research and Information in Education, Frankfurt am Main, Germany. ORCID: 0000-0001-9219-222X
[2] Email: g.ciordas-hertel@dipf.de Address: DIPF | Leibniz-Institute for Research and Information in Education, Frankfurt am Main, Germany. ORCID: 0000-0003-1589-7845
[3] Email: winter@psych.uni-frankfurt.de Address: Educational Psychology, Goethe-University, Frankfurt am Main, Germany. ORCID: 0000-0003-0350-1305
[4] Email: j.mordel@em.uni-frankfurt.de Address: Global Affairs Study and Teaching — Statistics Unit, Goethe-University, Frankfurt am Main, Germany. ORCID: 0000-0002-8469-1317
[5] Email: h.drachsler@psych.uni-frankfurt.de Address: DIPF | Leibniz-Institute for Research and Information in Education, Frankfurt am Main, Germany. ORCID: 0000-0001-8407-5314

1. Introduction

Personal digital devices are an essential part of learning in the 21st century. Their use for learning has become ubiquitous and indispensable, but they can also substantially interfere with learning goals. Problematic media use is a growing concern, particularly for some highly vulnerable populations (Sohn et al., 2019). This calls for increased research into the phenomenon and possible solutions.

Media use during learning can be described as the use of media for either on-task or off-task purposes (Liu et al., 2021; Wood & Zivcakova, 2015). On-task use aims to support the learning process. Examples include researching a topic, watching a lecture video, or collaborating online. On the other hand, digital media can be used in a variety of ways that having nothing to do with the task at hand: Playing a video game, using social media, or surfing news sites can distract learners by providing alternative, potentially more enjoyable activities. Of course, these activities are only a problem if they interfere with personal goals or well-being, which is indeed often the case: Media use that interrupts learning is widespread (Calderwood et al., 2014), is associated with worse academic performance (e.g., Jamet et al., 2020; Masood et al., 2020; Patil et al., 2019), and has negative effects on various markers of well-being and mental health (Demirci et al., 2015; Sohn et al., 2019). It, therefore, makes a difference to the interpretation of device usage data whether someone is learning or just enjoying their free time. During learning, it also makes a difference whether someone is using a device to support learning or to distract themselves. The ability to make these distinctions is necessary to answer several questions in the field of technology-enhanced learning. The mental costs of task-switching have been well documented in the fields of cognition and neuroscience (e.g., Fischer & Plesow, 2015; Jeong & Hwang, 2016; Parry & le Roux, 2019). However, our research aims to expand our understanding of why and how task-switching occurs (e.g., Liu et al., 2021; Dönmez & Akbulut, 2021; Baumgartner & Wiradhany, 2022) and its potential impact on the overall learning process.

Fundamental questions about the use of media for learning are still unresolved. For example, while there are laboratory studies that capture the extent and patterns of media use during learning (Calderwood et al., 2014; Rosen et al., 2013), there is a lack of objective data on how this behaviour occurs while learning at home (Wang et al., 2015). This includes questions about how often and how long learners switch to off-task content, which content is involved, how it affects learning outcomes, and if behaviour can be classified as modern (post-pandemic) learning strategies (Popławska et al., 2021; Beuckels et al., 2021). There is also the open question of when and how a switch to learning-relevant content can lead to learners losing sight of their actual goal and becoming engaged in off-task content. Such questions have direct relevance to course design considerations. For example, if it becomes apparent that links to external websites such as YouTube lead to a loss of engagement, then a stronger effort can be made to avoid this.

Another area where a distinction between on- and off-task use could contribute is digital self-control tools, which are apps and programs designed to prevent distractions from digital media by, for example, blocking distracting websites. These tools should only be active during learning, and they should not interfere with on-task media use for learning purposes, or during leisure time (Lyngs et al., 2020). The fact that most tools cannot make this distinction is a drawback to their adoption, and users stop using them even though they are actually helpful (Lyngs et al., 2022). It is an important and open question whether it is possible to develop tools that can make the distinction between off-task distractions and leisure time use (Biedermann et al., 2021).

To gain insight into the behaviours associated with on- or off-task media usage, the challenge is to record activities as accurately as possible, while at the same time knowing the motive behind each of them. Yet, almost all of what we know about media use during learning comes from studies that either use self-reports, which provide aggregations over longer periods of time, or from observations in the classroom (Jamet et al., 2020) or in a monitored learning environment (Calderwood et al., 2014; Liu et al., 2021). Both types of obtaining data about media use have severe limitations.

Observation studies are costly and time consuming, both for the researchers and for the participants. Furthermore, they lack external validity as they do not provide the possibility to observe behaviour in the field, for instance, learning at home. As a result, most studies use self-reporting, even though self-reporting of media use has been repeatedly shown to be highly inaccurate. A review and meta-analysis by Parry et al. (2021) found that self-reported and logged media use correlate only moderately. These results can be attributed to various aspects biasing the reliability of self-reported media use. For example, explicit awareness and memory are limited, and arguably particularly so for habitual behaviour (Sniehotta & Presseau, 2012). Thus, recalling all the apps and websites that one has used throughout a longer time span is particularly challenging. For instance, a self-report about media usage during one lecture (Jamet et al., 2020) covers a range of more than an hour, a time span during which typically dozens of different digital activities occur, potentially on many devices simultaneously (May & Elder, 2018).

Automatic logging of device activity can provide the means to avoid the limitations of direct observations and self-reports. However, there are challenges to this type of data collection, both technically and fundamentally. For example, when we are interested in the internal states of a learner — their goals and intentions that drive their media use, i.e., in the distinction between on-task and off-task behaviour — then logs of device use can prove insufficient as well.

Thus, to bridge the gap between the advantages of automatic logging (i.e., the accuracy) and self-reports (i.e., the way to obtain information about internal states), we present a system that uses automated logging and is enriched by user annotations. To delineate the need for the various components of the logging system, we begin by describing its requirements.

2. Requirements for Automatic Logging of On- and Off-Task Behaviour

A multitude of apps, browser extensions, and programs exist that log device usage data, some pre-installed on operating systems, to help users keep track of their usage times (Roffarello & De Russis, 2019). When using a logging application on a single device to collect research data, the resulting data set would have several limitations: 1) there would be no way of telling which part of the data was generated during learning; 2) there would be no information about the activities on other devices that a user might have used at the same time; 3) the data would not reveal why an activity was performed.

2.1 Accounting for the Learning Context

Logging should only be active in the context that is the subject of the research; otherwise we receive a lot of data that is of no interest. Even more problematic is that with such data, we also don't even know what data has been generated during learning in the first place. How can contextual logging be activated? The simplest way would be to let participants activate and deactivate the logging themselves. The glaring problem is that participants must remember to perform the activation and deactivation. If they forget to turn it on, there is no data. If they forget to turn it off, the data is no longer correctly contextualized. A more comfortable way would be via automatic context detection via sensors (Ciordas-Hertel et al., 2021). This is a promising avenue, but right now this is still early research that would also require that all participants own and wear mobile sensors like smartwatches. Another approach, and the one that we have chosen, is the activation through the detection of activity on learning materials in an online learning environment. Whenever a participant is active in the learning material, we can say with high confidence that they have an intention to learn. Once they stop their activity on the material for a longer time, we can infer that they have likely stopped learning. The caveat is that we are not able to tell whether they continued learning with materials that may not be part of the online learning environment, but wherever this activation is possible, we get data that is highly contextualized.

2.2 Logging Multiple Data Sources

Practically every college student owns a smartphone and many students also own a second device such as a notebook or tablet (Poll, 2015, p. 20). To gain insights into media behaviour, this circumstance must be considered. Imagine the situation where a student is studying and researching on a notebook while regularly distracting themselves with chats on their smartphone. By only logging activities on the notebook, we report on students who never distract themselves and only use media extremely diligently for learning-related activities. This leads to biases in the log data.

However, we must address the fact that logging is not equally possible on all types of devices. Especially in the walled gardens of mobile operating systems, capabilities are often restricted by the system vendors, and these restrictions also tend to change over time. Collecting usage data may be possible in one version of an operating system, but not on a later version. At this point in time, logging of activity in the browser, on computers, and on Android devices is possible, but on iOS devices only vendor-provided system apps can log app usage data automatically.

Ignoring iOS devices — which make up between a quarter and two-thirds of all devices — would introduce a significant bias in the data. To deal with this constraint, one can ask the users for "data donations" (Ohme et al., 2021). With data donations, participants voluntarily provide device logs to the researchers. We included this data collection approach in our logging system and fused the iOS data donations with the data from the other data sources.

In the study by Ohme et al. (2021), the authors report that data donations required technical skills from the participants, which sometimes proved limiting. While the concept of data donations admittedly adds friction to the process, it is currently the only way to obtain this data, and it also has the advantage of extending agency to the user while easing some of the privacy concerns. As we will argue next, a certain degree of interaction between the user and their logged data is necessary to be able to distinguish between on- or off-task device usage.

2.3 Annotation of Recorded Data

Many apps and websites can be used both for on- and off-task activities, but a simple measure of usage duration alone will be insufficient to distinguish why they were used. Websites such as YouTube contain entertainment, but one can also find a lot of high-quality learning content. Messenger apps are another example for dual use. They can be part of a learning process, for instance, when exchanging information with teachers or fellow students. On the other hand, they can be a source of distraction when messaging disrupts learning. In both cases, simply recording how often and for how long these technologies were used would not suffice to know whether the use should be considered on- or off-task.

To identify whether an activity was on-task or off-task, we can either try to log and analyze even more data, such as keyboard inputs, and use this data to try to infer more context. Or we can ask the subjects to review and annotate their data. The former, i.e., more data collection, would entail more privacy concerns and, even then, there would be no guarantee that the inference would be possible. Therefore, and despite the shortcomings of self-reports, we see it as necessary to include a means of asking the users about the logged device activities.

2.4 Privacy Considerations

In addition to the technical requirements, there are privacy considerations. Although not requirements per se, considering them is still necessary. An ever-increasing part of one's life is managed on digital devices, and highly private matters like dating activities or doctor's appointments are organized or carried out via digital devices. It is understandable that users are selective with whom they choose to share their data. One could argue that users share their data all the time, and with questionable parties for that matter. This appears to be true, and users on digital devices indeed often end up sharing much more data and information than they claim is acceptable (Kokolakis, 2017), a phenomenon known as the privacy paradox. However, this does not exempt researchers from moral and legal obligations, such as data minimization principles of the European General Data Protection Regulation. Further, the privacy paradox does not entail that privacy and trust are inconsequential, as the extensiveness of logging has an impact on acceptance and, ultimately, the willingness to participate (Drachsler & Greller, 2016; Lorenz et al., 2013; May & George, 2011).

3. Implementation of the Logging System

We addressed the challenges in the following ways: For the contextual activation, our system is connected to the Moodle learning management system (LMS), where activity triggers the activation of the logging system. To incorporate as many data sources as possible, we combined automatic logging with data donations. For the enrichment of log-data with self-reported data annotation, we created a user-facing website, which gives users the opportunity to review their own data. As a backend to combine and process the data, we build upon the "EduTex" system (Ciordas-Hertel et al., 2021). The entire system is available online under https://gitlab.com/edutex.

3.1 Accounting for the Learning Context

To address the context, the logging was activated whenever any interactions within the LMS were detected via a custom plugin that logged all in-browser events like scrolling, clicking, and mouse movements. The logging remained active for 10 minutes after the last interaction in the LMS was observed. If the user did not return to the LMS within that period, the data of this period was afterwards discarded as not belonging to the learning session. If the user did return within the 10 minutes, then the data was logged as during learning. We note that these 10 minutes are a bit arbitrary, and an investigation of the "real" break patterns can only occur after the fact. Indeed, investigating these kinds of break patterns is one of the reasons why such a logging system is necessary in the first place. Ultimately, we chose the duration of 10 minutes based on the fact that all videos in the course were shorter than 10 minutes. So, sitting perfectly still and watching the videos would not result in a session-break. Furthermore, 10 minutes could be communicated to the learners as a plausible break duration.

3.2 Logging of Multiple Data Sources and Data Donations

As data sources, we incorporated the browser, android devices, and iOS devices. Where possible, we used automatically triggered logging. For iOS, we used data donations.

In the browser, we implemented logging via a browser extension for Chrome, Firefox, and Edge browsers. The browser extension logged the domain of the currently visited website, while discarding any additional information from the URL (i.e., it only logged "example.com" instead of "example.com/user/johndoe"). This was done to reduce privacy concerns because other parts of the URL, like the path and parameters, can contain sensitive and personally identifying information. The browser extension was distributed through the extension store (Chrome, Edge) and via a download page with instructions (Firefox).

In the android app, a background service regularly polled the backend to check if it should be active according to the activation rule. While it was active, the app logged the package name for each app that a user opened. To access this data, the background service used the "UsageStatsManager" API. The app was distributed through the Google Play store.

To obtain iOS usage data, we used the concept of a data donation by asking users to export and upload their devices "privacy report" files. This file contains app usage data, has exact timestamps, and can thus be contextualized to the learning activity. We made use of this feature and created an upload functionality where our users can upload their file. The backend system retrieved the app identifier and timestamps from the app activity report files, and retroactively contextualized the app activity data to the learning session with the LMS log data from the same participant.

3.3 The Annotation Page

We created the annotation page as a separate website where participants could review and annotate their own activity during learning (see the screenshot in Figure 1). The display of activities was set up so that participants could see one session — defined as an activity with no break longer than 10 minutes — at a time. For each activity, participants were asked to annotate which need they wanted to fulfill with this activity. The answer options from which the participants could choose were “research/information,” “relaxation,” “entertainment,” “social interactions,” “habitual behaviour,” “I don’t want to answer,” and “other.” Following a uses-and-gratification-approach, answers hereby attempt to explain actions either through the satisfaction of needs or hoped-for benefits (Lin et al., 2014). Multiple selections were possible. For further contextualization of the activities, the LMS interactions of the participant were also displayed on the same page. The rationale was to support participant recall, so that they could better remember what they were doing and why they chose to initiate an activity. For instance, they might remember that they had to look up something when they were reading a text, or they might remember that a video was very boring and so they started chatting.

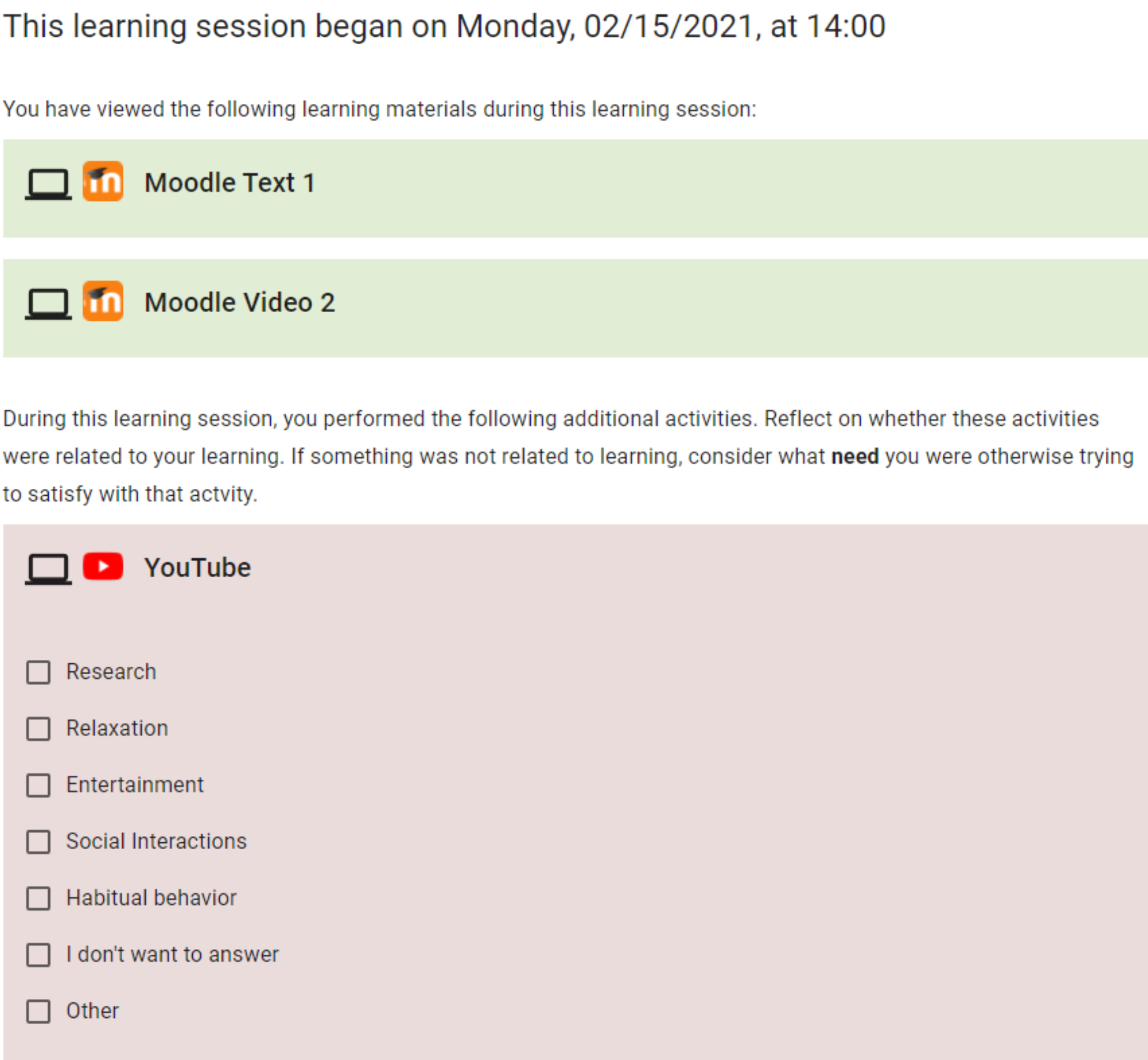


Figure 1. Screenshot of the annotation page that was part of a dashboard at the end of the course.

3.4 Backend System

The backend system stored, fused, and filtered the various data sources, and managed the activation of the logging where applicable. Upon request from the annotation page, the backend calculated sessions according to the activity in the LMS. All the activity within that time frame was processed as belonging to that session. Apps and web pages like empty tabs, launchers, or system processes were filtered out at this stage because they are only intermediate or sub-activities to the actual activities. The backend also performed data processing for presentation purposes. The logs originally contained package names (e.g., "com.google.android.gm" instead of "Gmail") or the domain of a website. The backend tried to obtain user-friendly display names and icons for each activity before showing them to the user. This information was not obtainable directly from the logs, so we used a mixture of public APIs (for websites and iOS apps) and commercial APIs (for Android apps) to obtain this information on the fly.

4. Formative Study Using the Logging System

We deployed the system as part of a study on the use of trace data in learning. We use results from this study as indicators to examine how our assumptions regarding the contextualization, the use of multiple devices, and the annotation of data are supported.

4.1 Methods

The study took place in the last unit of an asynchronous online course for student teachers over a full semester. The topic of the full course was the use of digital media in teaching, and the topic of the last unit was learning analytics. The whole course was ungraded, but students received ECTS for completing it. The passing criterion was that participants had to answer several questions at the end of each course unit. Every learning unit was available for 14 days except for one that took place during winter break (28 days). In the last two weeks, where the study took place, the students were instructed to observe their own activities during learning. They could do so by either using the logging system, or by self-recording their activities in a spreadsheet. Written instructions for setting up the logging on their devices of choice were given at the beginning of the course. The learning materials consisted of two texts and two videos. After completing the learning materials in the course, they had access to the dashboard. Use of the dashboard was suggested, but not mandatory to complete the course. The students were asked for consent on whether their data could be used for research purposes, which was not mandatory. Approval was obtained from the ethics committees of the participating institutions. In total, 388 students were enrolled in the course for these two final weeks. Out of those, 297 consented to have their data used for scientific analysis.

4.2 Results

First, we investigated participation for the various data sources. In total, 177 out of the 297 participants gave their consent and recorded any activity on any data source. To examine the extent to which the integration of different data sources played a role, we look at which data sources recorded activities. The most recorded data source was the browser (119), followed by iOS (75) and Android (26). The majority (135) recorded activity only on one data source, while 41 participants recorded the activity of two data sources, and one participant recorded activity on three different data sources. Sixty-two participants decided to use the self-report, which we are not analyzing here.

Regarding the contextual activation, we reviewed the activities that the system recorded during the learning sessions. Across all data sources of all consenting participants, 1387 activities were recorded. On average, each participant had 8.11 activities (SD = 9.10) recorded. From the recorded activities, 1031 were annotated by 163 participants (92%). We cannot know if there were activities that were incorrectly contextualized, but neither in the support forum nor in direct messages to the course administrators was there any mention of such an issue.

As an indicator for privacy concerns, we looked at the annotations where a participant chose “Don’t want to answer.” This option was chosen for 34 activities from 17 different participants. We checked whether those participants simply labelled all their activities as “Don’t want to answer.” This was not the case, and only one of the participants annotated all activities with “Don’t want to answer,” suggesting that participants made this choice very deliberately.

To explore if device usage includes both on-task and off-task purposes, we counted how often participants selected the annotation “Research.” This was the most common annotation overall (n = 514, 49.85%), especially in the browser (n = 329, 79.85%). For the other data sources, the annotations were more evenly distributed (see Figure 2). The browser results show a limitation of our logging approach: While the browser extension was capable of recording individual websites, this was not possible in the case of mobile browser apps. That is, whenever someone used the browser, the system logged only that the browser app was used in general, but not which website was accessed. Thus, the level of detail regarding the websites that participants accessed was lower on mobile devices than on the browser extension.

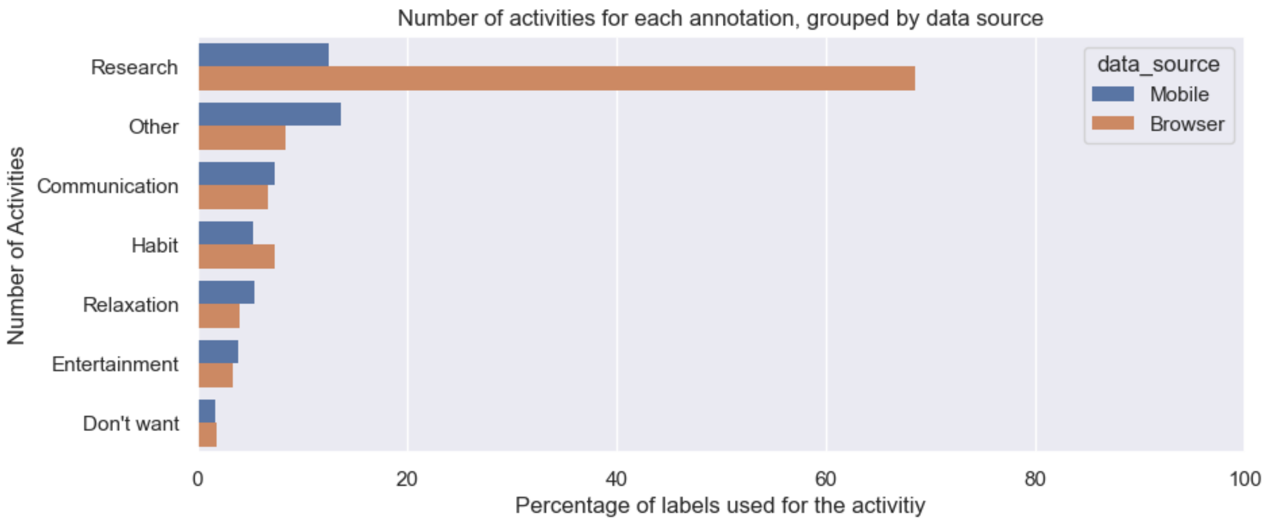


Figure 2. Percentage of the category selected as annotation, differentiated between browsers and mobile devices.

We also checked how often the participants annotated the same activity as used both for on-task and off-task purposes. For on-task activities, we used only the “research” annotation, and for off-task all annotations that were not “research,” “don’t want to answer,” or “other.” We grouped the activities into the categories “browser,” “search engine,” “email,” “video platform,” “image platform,” “messenger,” “note taking,” “social media,” and “other” (see Figure 3). In the “other” group were apps that we observed only once or twice in total. The results show that 24 different activities share both on- and off-task labels. The ratios differ strongly; for instance, the participants annotated browser apps in only 7% of the cases as off-task, while messenger apps and websites were annotated as off-task in 81% of the cases.

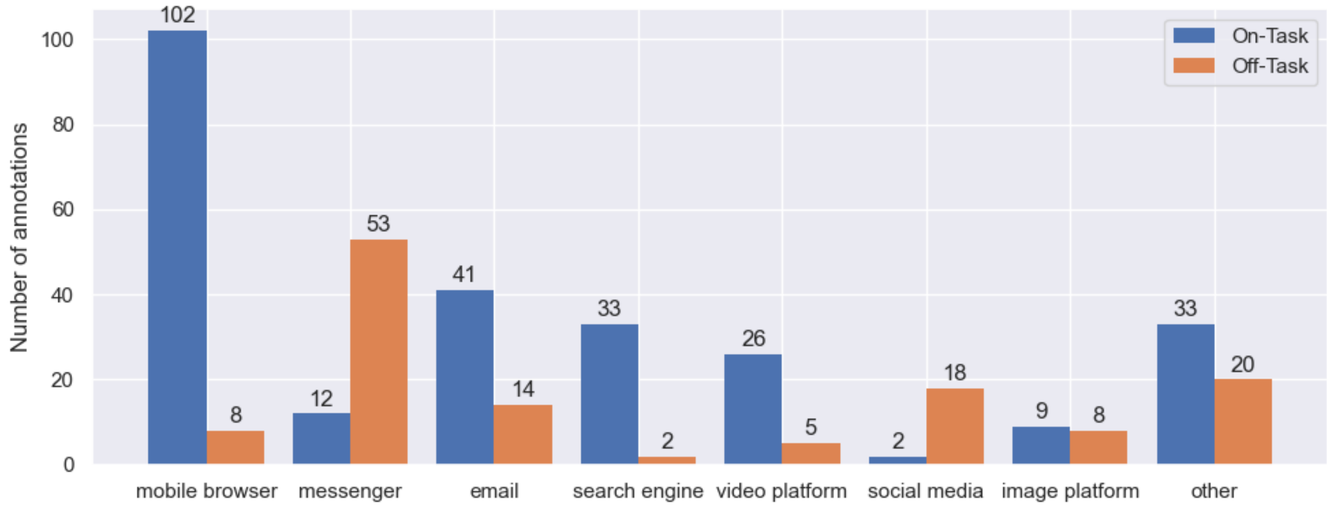


Figure 3. Categories of digital activities annotated by users as both on-task and off-task.

We know that self-report data becomes less accurate the longer the time between the event and the report. Thus, we looked at the time between the activity and the annotation of the activity. On average, the time difference was 12 hours and 58 minutes (SD = 39 hours and 31 minutes). A small number of outliers skew this time difference, with a maximum of 292 hours and 31 minutes between activity and annotation. Looking at percentiles, we see that most annotations were much closer to the activity: 50% of the activities were annotated less than 50 minutes after their enactment, 75% of the activities in 1 hour and 38 minutes or less, and 99% in 20 hours and 31 minutes or less.

5. Discussion

We argued that the logging of digital activities is only capable of distinguishing between on- and off-task activities during learning if the following requirements are met. The logging system must be active in the context of the learning, and it must be aware of the times where learning starts, and where it ends. Furthermore, logging must be possible from multiple devices at the same time, to account for the fact that learners today usually own and use more than one device. Finally, the system must be able to distinguish between on- and off-task activities. The results of our study support all these assumptions, and they show that our system can capture these differences.

5.1 Participation Across Data Sources

The high variation of activities across data sources underlines that capturing only a single data source would be insufficient. In our dataset, only a few people logged activities on more than one device. Since we did not ask for reasons for not choosing more than one data source, we can only guess the reasons. In some cases, it might be that there was simply no activity on a second device. However, it is also very likely that many participants were simply not willing to put in the effort required (Jürgens et al., 2020; Makhortykh et al., 2022). Interestingly, the participation rate of iOS users was more than double than that of android users.

This indicates either a highly skewed sample or that the friction of using this external app was higher than it was for the data donation (Ohme et al., 2021). Another observation regarding participation rates was that the highest rate was with the browser extension, where the installation was very fast and simple — just two clicks and no login. Another possible interpretation is that participants perceived the content of their smartphone to be more privacy sensitive. An investigation into these reasons remains an important future step because, as we highlighted, some inference on the data is only possible when the activity is logged across all user devices.

5.2 Annotations and the Distinction Between On- and Off-Task Activities

The annotations showed that both on- and off-task activities occur during learning; in several cases, the same digital activity was used in both contexts. The data annotation is also the part of the system with the most leeway and opportunities for adjustments to different research questions. In the formative study, the goal was to evaluate a uses and gratification approach, and the labels that participants could choose from reflected this. Studies with other research questions should consequently offer other labels as choices. This also goes hand in hand with how much information is presented to users. While in our study, several activities were summarized on one page (for example, displaying only "Wikipedia" instead of listing all Wikipedia pages accessed), in other studies it might be necessary to list the information in more detail. Of course, such adaptations also result in new requirements for participants.

Without the user annotations, the distinction into on- and off-task activities would hardly be possible. We therefore see the step of data annotation as unavoidable, but we must nevertheless note that they demand a great deal of effort from the learners. This consideration comes on top of the general ethical considerations surrounding the collection of data in learning environments (Drachsler & Greller, 2016; Slade & Prinsloo, 2013). This is a difficult dilemma to resolve, and we would like to emphasize above all that we consider transparency and voluntariness to be essential. We as researchers should always consider how we can minimize negative consequences for participants.

5.3 Contextual Activation

The contextual activation worked well but relies on the availability of learning materials in an online learning environment and the modification of that environment to trigger activation. It is also important to consider how to obtain contextualized data when the learning content is not available in a modifiable LMS such as Moodle, and accordingly one does not have an automatic trigger available. If automatic activation through an online learning environment is not possible, manual activation should be the preferred method. This would leave learners in charge of starting the logging process, which can lead to biases due to interpersonal differences, such as between more and less successful self-regulated learners. In future, the differences between manual and automatic activation should be evaluated. In this context, there are also promising research approaches to automatic context recognition that could trigger logging in the future; future studies could evaluate how well these methods are suited to the task (Ciordas-Hertel et al., 2021; Laput & Harrison, 2019).

5.4 Potential for Educational Technologies and Learning Analytics

By combining LMS trace data with device activity data, we can gain insights that may not be possible with LMS data alone. If we consider that any activity on digital devices leads to a break in interactions in the LMS, we can see that our system could help to shed more light on inactivity behaviour in general. This could, for instance, help to further improve the calculation of the highly relevant time-on-task indicator. Time-on-task can be calculated differently depending on how sequences of inactivity are interpreted (Kovanović et al., 2015; Leinonen et al., 2022) but breaks in the interactions still present a challenge. A break in the activity stream could mean that a student was taking notes (i.e., more time spent engaged with the task) or that a student was using their smartphone (i.e., less time spent engaged with the task). With our data, one could investigate these break patterns and try to determine when a sequence of inactivity should actually mean more time-on-task, and when it should be subtracted.

Looking at interaction data in the context of other media activities also allows us to look for direct connections between the two. We could, for example, examine whether there are click- or scroll-level interaction patterns in the LMS that indicate an imminent (off-task) disengagement on a digital device. These relationships between user interaction patterns and states of disengagement have been studied previously (e.g., Dias da Silva & Postma, 2020; Thorpe et al., 2022), and experimental studies have shown that switching between media is heralded several seconds in advance by physiological signals such as skin conductance (Yeykelis et al., 2014). As such, one could investigate if and how interactions like mouse movements or scrolling patterns in an LMS can also predict switching to off-task activities. If it is possible to find interaction patterns that temporally precede off-task behaviour, one could conceive early interventions that might help learners stay on-task, for example by warning them of their loss of attention early on. It is not clear whether something like this can be implemented in practice, as it depends on whether this temporal prediction is possible at all. However, there are also practical implementations that we consider directly feasible.

Temporal prediction, however, might not even be necessary to improve already existing digital self-control tools like website blockers. The learning activity triggers could be directly linked to and activated by these tools only during learning, thus avoiding the shortcomings of current tools, which do not distinguish between learning and leisure activities. Without this distinction, users must micromanage their self-control tools, and turn them on and off, depending on whether they are learning or not. Human habits, conveniences, and forgetfulness can then lead to abandonment of any self-control tool altogether. With automatic activation based on learning activity, this burden could be removed.

5.5 Limitations of the Data and Other Opportunities for Future Work

The use of our system resulted in a very rich and contextualized dataset, but there is still room for improvement. The annotations were not always temporally close to the logged behaviour and thus the risk of inaccurate recall remained (Schwarz, 2007; Tourangeau, 2000). Furthermore, despite our efforts to improve the self-report annotations through the contextualized activities, we must assume that weaknesses of self-reports are still present, such as social desirability and biased post-hoc attributions. We have no way of investigating how prominent these were in our dataset, but there is little reason to assume that they would not be present. Future studies could experiment with asking participants more frequently to self-report their annotations, e.g., by making use of experience sampling. In general, we relied on proxy indicators for on- versus off-task activities. Especially the "research" annotation is subject to interpretation. For instance, participants could have interpreted it as "researching what movie to watch later." We still expect that participants interpreted the annotations mostly as intended, but we cannot be completely certain about that.

In our study, we failed to ask students explicit questions about the acceptability of tracking and annotation efforts. For the sake of the validity and reliability of our proposed method, this important aspect should be given more attention in future studies.

We have, thus far, also not included logging for desktop applications. This is important especially for games, which typically run as standalone applications on the computer. There is no technical obstacle to that, and it could be included in the system in future.

6. Conclusion

In this paper, we presented a rationale and a system design for logging of device activity in educational settings to allow for the differentiation between on- and off-task activities. Related work shows that media use can have serious adverse consequences, but at the same time, the methodology for researching the related phenomena is often inadequate. Self-reports lack accuracy, and lab studies lack external validity. Consequently, device use must be logged in order to draw valid inferences, and the design of the logging system needs to implement several key functionalities.

This requires a much higher technical effort for those who want to conduct similar research. It is much easier to issue a questionnaire than to create and distribute a logging system. However, the already known shortcomings of self-report questionnaires combined with our findings demonstrate the need for thoughtful collection of media use data. This is necessary because understanding the positive and negative effects of media use is a key area of research whose validity has suffered from inadequate methods.

With our logging system, we have presented an approach to improve data quality by enriching log data with user annotations. Our system can capture data in a way that provides insights into device use during learning that would not otherwise be possible. With improved datasets, both basic research and concrete applications such as self-control tools can be advanced.

7. Declaration of Conflicting Interest

The authors declared no potential conflicts of interest with respect to the research,

authorship, and/or publication of this article.

8. Funding

The authors declared no financial support for the research, authorship, and/or publication of this article.

References

Beuckels, E., Ye, G., Hudders, L., & Cauberghe, V. (2021). Media multitasking: A bibliometric approach and literature review. *Frontiers in Psychology*, *12*, 623643. https://doi.org/10.3389/fpsyg.2021.623643

Baumgartner, S. E., & Wiradhany, W. (2022). Not all media multitasking is the same: The frequency of media multitasking depends on cognitive and affective characteristics of media combinations. *Psychology of Popular Media*, *11*(1), 1–12. https://doi.org/10.1037/ppm0000338

Biedermann, D., Schneider, J., & Drachsler, H. (2021). Digital self-control interventions for distracting media multitasking: A systematic review. *Journal of Computer Assisted Learning*, *37*(5), 1217–1231. https://doi.org/10.1111/jcal.12581

Calderwood, C., Ackerman, P. L., & Conklin, E. M. (2014). What else do college students "do" while studying? An investigation of multitasking. *Computers & Education*, *75*, 19–29. https://doi.org/10.1016/j.compedu.2014.02.004

Ciordas-Hertel, G.-P., Rödling, S., Schneider, J., Di Mitri, D., Weidlich, J., & Drachsler, H. (2021). Mobile sensing with smart wearables of the physical context of distance learning students to consider its effects on learning. *Sensors*, *21*(19), 6649. https://doi.org/10.3390/s21196649

Demirci, K., Akgönül, M., & Akpinar, A. (2015). Relationship of smartphone use severity with sleep quality, depression, and anxiety in university students. *Journal of Behavioral Addictions*, *4*(2), 85–92. https://doi.org/10.1556/2006.4.2015.010

Dias da Silva, M. R., & Postma, M. (2020). Wandering minds, wandering mice: Computer mouse tracking as a method to detect mind wandering. *Computers in Human Behavior*, *112*, 106453. https://doi.org/10.1016/j.chb.2020.106453

Dönmez, O., & Akbulut, Y. (2021). Timing and relevance of secondary tasks impact multitasking performance. *Computers and Education, 161*, 104078D. https://doi.org/10.1016/j.compedu.2020.104078

Drachsler, H., & Greller, W. (2016). Privacy and analytics: It's a DELICATE issue a checklist for trusted learning analytics. *Proceedings of the 6th International Conference on Learning Analytics and Knowledge* (LAK '16), 25–29 April 2016, Edinburgh, UK (pp. 89–98). ACM Press. https://doi.org/10.1145/2883851.2883893

Fischer, R., & Plessow, F. (2015). Efficient multitasking: Parallel versus serial processing of multiple tasks. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.01366

Jamet, E., Gonthier, C., Cojean, S., Colliot, T., & Erhel, S. (2020). Does multitasking in the classroom affect learning outcomes? A naturalistic study. *Computers in Human Behavior*, *106*, 106264. https://doi.org/10.1016/j.chb.2020.106264

Jeong, S. H., & Hwang, Y. (2016). Media multitasking effects on cognitive vs. attitudinal outcomes: A meta-analysis. *Human Communication Research, 42*. https://doi.org/10.1111/hcre.12089

Jürgens, P., Stark, B., & Magin, M. (2020). Two half-truths make a whole? On bias in self-reports and tracking data. *Social Science Computer Review*, *38*(5), 600–615. https://doi.org/10.1177/0894439319831643

Kokolakis, S. (2017). Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & Security*, *64*, 122–134. https://doi.org/10.1016/j.cose.2015.07.002

Kovanović, V., Gašević, D., Dawson, S., Joksimović, S., Baker, R. S., & Hatala, M. (2015). Penetrating the black box of time-on-task estimation. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 184–193). ACM Press. https://doi.org/10.1145/2723576.2723623

Laput, G., & Harrison, C. (2019). Sensing fine-grained hand activity with smartwatches. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (CHI '19), 4–9 May 2019, Glasgow, Scotland, UK (Paper No. 338). ACM Press. https://doi.org/10.1145/3290605.3300568

Leinonen, J., Castro, F. E. V., & Hellas, A. (2022). Time-on-task metrics for predicting performance. *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education* (SIGCSE '22), 3–5 March 2022, Providence, RI, USA (pp. 871–877). ACM Press. https://doi.org/10.1145/3478431.3499359

Lin, Y.-H., Fang, C.-H., & Hsu, C.-L. (2014). Determining uses and gratifications for mobile phone apps. In J. J. Park, Y. Pan, C.-S. Kim, & Y. Yang (Eds.), *Future information technology* (pp. 661–668). Springer. https://doi.org/10.1007/978-3-642-55038-6_103

Liu, Y., Deng, L., Lin, L., & Gu, X. (2021). Patterns of triggers for on-task and off-task behaviors: University students in independent study. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2021.1905003

Lorenz, B., Sousa, S., & Tomberg, V. (2013). Privacy awareness of students and its impact on online learning participation: A case study. *Proceedings of the IFIP WG 3.4 International Conference* (OST 2012), 20 July–3 August 2012, Tallinn, Estonia. Springer. https://doi.org/10.1007/978-3-642-37285-8_21

Lyngs, U., Lukoff, K., Csuka, L., Slovák, P., Van Kleek, M., & Shadbolt, N. (2022). The Goldilocks level of support: Using user reviews, ratings, and installation numbers to investigate digital self-control tools. *International Journal of Human–Computer Studies*, *166*, 102869. https://doi.org/10.1016/j.ijhcs.2022.102869

Lyngs, U., Lukoff, K., Slovak, P., Seymour, W., Webb, H., Jirotka, M., Zhao, J., Van Kleek, M., & Shadbolt, N. (2020). "I just want to hack myself to not get distracted": Evaluating design interventions for self-control on Facebook. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (CHI '20), 25–30 April 2020, Honolulu, HI, USA (pp. 1–15). ACM Press. https://doi.org/10.1145/3313831.3376672

Makhortykh, M., Urman, A., Gil-Lopez, T., & Ulloa, R. (2022). To track or not to track: Examining perceptions of online tracking for information behavior research. *Internet Research, 32*(7), 260–279. https://doi.org/10.1108/INTR-01-2021-0074

Masood, A., Luqman, A., Feng, Y., & Ali, A. (2020). Adverse consequences of excessive social networking site use on academic performance: Explaining underlying mechanism from stress perspective. *Computers in Human Behavior*, *113*, 106476. https://doi.org/10.1016/j.chb.2020.106476

May, K. E., & Elder, A. D. (2018). Efficient, helpful, or distracting? A literature review of media multitasking in relation to academic performance. *International Journal of*

Educational Technology in Higher Education, *15*, 13. https://doi.org/10.1186/s41239-018-0096-z

May, M., & George, S. (2011). Privacy concerns in e-learning: Is using tracking system a threat? *International Journal of Information and Education Technology*, *1*(1), 1–8. https://doi.org/10.7763/IJIET.2011.V1.1

Ohme, J., Araujo, T., de Vreese, C. H., & Piotrowski, J. T. (2021). Mobile data donations: Assessing self-report accuracy and sample biases with the iOS Screen Time function. *Mobile Media & Communication*, *9*(2), 293–313. https://doi.org/10.1177/2050157920959106

Parry, D. A., Davidson, B. I., Sewall, C. J. R., Fisher, J. T., Mieczkowski, H., & Quintana, D. S. (2021). A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. *Nature Human Behaviour*, *5*, 1535–1547. https://doi.org/10.1038/s41562-021-01117-5

Parry, D. A., & le Roux, D. B. (2019). Media multitasking and cognitive control: A systematic review of interventions. *Computers in Human Behavior, 92*, 316–327. https://doi.org/10.1016/j.chb.2018.11.031

Patil, R., Brown, M., Ibrahim, M., Myers, J. L., Brown, K., Khan, M., & Callaway, R. (2019). Digital distraction outside the classroom: An empirical study. *Journal of Computing Sciences in Colleges*, *34*(7), 46–55.

Poll, H. (2015). *Pearson student mobile device survey 2015: College students.* Pearson.

Popławska, A., Szumowska, E., & Kuś, J. (2021). Why do we need media multitasking? A self-regulatory perspective. *Frontiers in Psychology, 12*, 624649–624649. https://doi.org/0.3389/fpsyg.2021.624649

Roffarello, A. M., & De Russis, L. (2019). Towards detecting and mitigating smartphone habits. *Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers* (UbiComp/ISWC '19), 9–13 September 2019, London, United Kingdom (pp. 149–152). ACM Press. https://doi.org/10.1145/3341162.3343770

Rosen, L. D., Carrier, L. M., & Cheever, N. A. (2013). Facebook and texting made me do it: Media-induced task-switching while studying. *Computers in Human Behavior*, *29*(3), 948–958. https://doi.org/10.1016/j.chb.2012.12.001

Schwarz, N. (2007). Retrospective and concurrent self-reports: The rationale for real-time data capture. In A. A. Stone, S. Shiffman, A. A. Atienza, & L. Nebeling (Eds.), *The science of real-time data capture: Self-reports in health research* (pp. 11–26). Oxford University Press.

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*, *57*(10), 1510–1529. https://doi.org/10.1177/0002764213479366

Sniehotta, F. F., & Presseau, J. (2012). The habitual use of the self-report habit index. *Annals of Behavioral Medicine*, *43*(1), 139–140. https://doi.org/10.1007/s12160-011-9305-x

Sohn, S. Y., Rees, P., Wildridge, B., Kalk, N. J., & Carter, B. (2019). Prevalence of problematic smartphone usage and associated mental health outcomes amongst children and young people: A systematic review, meta-analysis and GRADE of the evidence. *BMC Psychiatry*, *19*, 356. https://doi.org/10.1186/s12888-019-2350-x

Thorpe, A., Friedman, J., Evans, S., Nesbitt, K., & Eidels, A. (2022). Mouse movement trajectories as an indicator of cognitive workload. *International Journal of Human–Computer Interaction*, *38*(15), 1464–1479. https://doi.org/10.1080/10447318.2021.2002054

Tourangeau, R. (2000). Remembering what happened: Memory errors and survey reports. In A. A. Stone, J. S. Turkkan, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman, & V. S. Cain (Eds.), *The science of self-report: Implications for research and practice* (pp. 29–47). Lawrence Erlbaum Associates Publishers.

Wang, Z., Irwin, M., Cooper, C., & Srivastava, J. (2015). Multidimensions of media multitasking and adaptive media selection. *Human Communications Research, 41*, 102–127. https://doi.org/10.1111/hcre.12042

Wood, E., & Zivcakova, L. (2015). Understanding multimedia multitasking in educational settings. In L. D. Rosen, N. A. Cheever, & L. M. Carrier (Eds.), *The Wiley Handbook of Psychology, Technology, and Society* (pp. 404–419). John Wiley & Sons Ltd. https://doi.org/10.1002/9781118771952.ch23

Yeykelis, L., Cummings, J. J., & Reeves, B. (2014). Multitasking on a single device: Arousal and the frequency, anticipation, and prediction of switching between media content on a computer. *Journal of Communication*, *64*(1), 167–192. https://doi.org/10.1111/jcom.12070

Paper 2

Biedermann, D., Schneider, J., Ciordas-Hertel, G.-P., Eichmann, B., Hahnel, C., Goldhammer, F., & Drachsler, H. (2023). *Detecting the Disengaged Reader—Using Scrolling Data to Predict Disengagement during Reading*. LAK23: 13th International Learning Analytics and Knowledge Conference, 585–591. https://doi.org/10.1145/3576050.3576078

Detecting the Disengaged Reader - Using Scrolling Data to Predict Disengagement during Reading

DANIEL BIEDERMANN, DIPF | Leibniz Institute for Research and Information in Education, Germany
JAN SCHNEIDER, DIPF | Leibniz Institute for Research and Information in Education, Germany
GEORGE-PETRU CIORDAS-HERTEL, DIPF | Leibniz Institute for Research and Information in Education, Germany
BEATE EICHMANN, DIPF | Leibniz Institute for Research and Information in Education, Germany
CAROLIN HAHNEL, DIPF | Leibniz Institute for Research and Information in Education, Germany
FRANK GOLDHAMMER, DIPF | Leibniz Institute for Research and Information in Education, Germany
HENDRIK DRACHSLER, DIPF | Leibniz Institute for Research and Information in Education, Germany

When reading long and complex texts, students may disengage and miss out on relevant content. In order to prevent disengaged behavior or to counteract it by means of an intervention, it is ideally detected an early stage. In this paper, we present a method for early disengagement detection that relies only on the classification of scrolling data. The presented method transforms scrolling data into a time series representation, where each point of the series represents the vertical position of the viewport in the text document. This time series representation is then classified using time series classification algorithms. We evaluated the method on a dataset of 565 university students reading eight different texts. We compared the algorithm performance with different time series lengths, data sampling strategies, the texts that make up the training data, and classification algorithms. The method can classify disengagement early with up to 70% accuracy. However, we also observe differences in the performance depending on which of the texts are included in the training dataset. We discuss our results and propose several possible improvements to enhance the method.

ACM Reference Format:

Daniel Biedermann, Jan Schneider, George-Petru Ciordas-Hertel, Beate Eichmann, Carolin Hahnel, Frank Goldhammer, and Hendrik Drachsler. 2023. Detecting the Disengaged Reader - Using Scrolling Data to Predict Disengagement during Reading. In *LAK23: 13th International Learning Analytics and Knowledge Conference (LAK 2023), March 13–17, 2023, Arlington, TX, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3576050.3576078

1 INTRODUCTION

Reading is a central cultural technique in society that enables students to learn from written information. However, students will not always read through a text that is part of their curriculum completely. Especially in online learning scenarios in higher education, learners have a lot of leeway and can work on texts in a self-directed manner. Still, they may abandon a text for various reason even though they intended to work on it. Students might get distracted by a notification from their smartphone or start mind wandering and scrolling through the material without actually reading it [8]. Consequently, if they do not manage to return to the text and re-engage with it, they will miss out on course content. Accordingly, non-completion of learning materials can be considered as a form of behavioral disengagement.

We use the term disengagement in this paper, whenever we refer to the behavioral disengagement that is observable when a student is not completing a reading assignment [5].

Reading challenging texts is an integral part of many curricula [22]. However, the risk of disengagement increases with text length and difficulty [9]. Since it is not realistic or purposeful to condense all curricular content to tweet size, we need to do the best we can to support students in meeting reading-related learning challenges. Predicting if and when a student will struggle to complete their reading could facilitate the development of targeted interventions for both learners and instructors. For instance, if digital distractions are the culprit, the prediction of disengagement could improve tools that restrict access to distracting content [2].

However, preventive interventions are only effective if disengaged behavior can be predicted individually for each time a student reads a text at an early stage. Moreover, the prediction should be possible based purely on easily generated observational data, such as trace data. Otherwise, it is hardly applicable in practice. Previous work has shown that the potential exists. In this paper, we contribute to the prediction of disengagement by presenting a method for the temporal prediction of disengagement based on scrolling data.

1.1 Related Work

There have been attempts to predict behavior related to behavioral disengagement based on trace data with machine learning models. The models classify whether a set of interactions belongs to the class of those students who completely read a text or not. In the studies by Cocea and Weibelzahl [6] and Mills et al. [17], the text content was distributed across several pages and the classification models used this information as an essential feature. The page-based approach could predict the completion of the latter pages from interactions with earlier pages with above-chance accuracy. However, this approach is only applicable if a text is presented on several pages. Especially in online settings, texts often must be scrolled from top to bottom because they do not fit on a single screen, as it is the case in our data set. Consequently, models which rely on a measurable page transition cannot be applied.

An almost universally available feature is the time-on-task on the text [6, 18]. For our study, we define time-on-task as the time it takes for a student to work on a text. The rationale behind using time-on-task is that there is an expected reading time for a text based on its characteristics (e.g., its length and difficulty). Accordingly, more complex texts should result in more time that the reader spends on them [10]. Using time-on-task as an indicator for text completion has the advantage that it can be derived purely from trace data. Human observation is not required. However, it can still be subject to distortions, such as excessive idle time on texts, which can artificially inflate time-on-task [12]. Imagine a student opening their text, then going away for 15 minutes. The time on the text will indicate that they spent 15 minutes on their reading assignment, while it was actually much less. Another crucial disadvantage is that time-on-task can only be utilized after the reader has stopped interacting with the task and can thus not be used to predict the behavior.

1.2 This Study

In this study, we take an approach that relies purely on trace data captured during the reading task to classify whether early interactions will result in disengagement. Our method aims to achieve this classification while the learner is still working on the document that contains the text. We utilize scrolling trace data as the feature to predict disengagement. When navigating on a document larger than the available space on the screen, scrolling is necessary to change the position of the currently visible area. This visible area is called the viewport. If you read this text on a digital device that is not large enough to display the whole text at once, there is a reasonable chance that you scroll your viewport through the document in order to read everything.

From the viewport data, it is possible to infer user attention on a document with comparable performance to the use of eye-tracking [11, 13, 14, 20]. This can be achieved by using the viewport scrolling data to infer the time spent on individual parts of the document instead of just capturing the time a student spends on the entire document as it is done in time-on-task approaches. However, with both time-on-task and viewport approaches, we cannot determine whether students were reading attentively and processing the text. But we can determine whether individual sections of the text were visible on the screen at least long enough to make in-depth reading possible. We use this potential and derive an indicator that tells us whether a reader is likely to completely have read a text. Similar to time-on-task, this indicator is available only after the reader has left the text, and thus we use it only as the class label for the classification task.

The actual data we classify is the vertical scrolling position of the viewport. We interpret the viewport's vertical position as a function of time and thus convert it into a time series representation (see figure 1). This representation can be obtained at any time and is continuously updated while the reader still interacts with the text. Compared to global features (e.g., the average scrolling acceleration across the whole interaction), the time series representation preserves temporal relationships and patterns in the data.

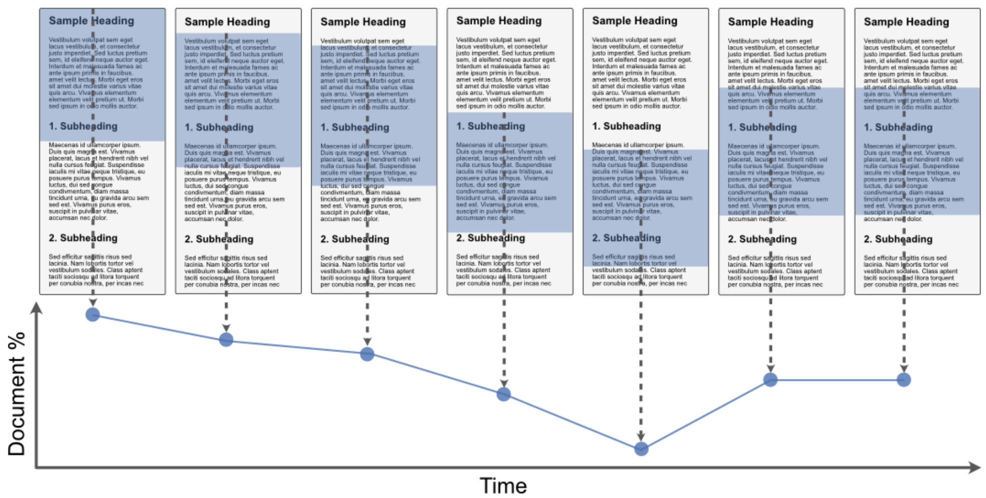


Fig. 1. Illustration of how viewport data is transformed into its time series representation. At fixed intervals, the position of the top of the viewport is sampled in relation to the document.

1.3 Time Series Classification

Time series classification differs from traditional classification tasks because the features in time series are often highly correlated, and the ordering of features is relevant. Numerous algorithms have been developed and refined for this task. They perform differently depending on the discriminatory features they use to assign a class to a time series. For example, the discriminatory features might be equally distributed over the whole time series, repeating or not, and so forth. The review article of Bagnall et al. [1] reports details regarding the taxonomy of classification algorithms.

To the best of our knowledge, there has been no prior work that attempts to classify time series representations of scrolling data. Thus, this study seeks to show new ways to make use of this data by exploring several aspects of this approach and evaluating the method on a dataset from a natural online learning setting.

In the following sections, we first describe the method of converting scrolling data to its time series representation and deriving the disengagement criterion. Next, we evaluate how our method performs on our dataset. We describe the manipulation of several aspects of the data and the algorithm selection to identify the strengths and weaknesses of our method.

2 DATASET AND PREPROCESSING

2.1 Dataset

The data for the study was acquired from a semester-long, fully online Moodle course from two German universities. The course topic was the use of digital media in education, and the course was designed for pre-service teachers. It consisted of five course units, each one available for two weeks. Every course unit contained several learning materials and could be accessed by the students at any time during the respective time of availability. There was a total of 16 texts and five videos as learning materials in the course. The students could complete the materials of each unit in the order of their choosing. The course was not graded, but students received credit for full participation. We used the data of those 565 students who fully completed the course for our analyses.

2.2 Data Tracking

The data were collected with a Moodle plugin logging mouse and keyboard interactions from each participant during the course. Among many other interactions, the plugin also tracked the viewport position, width, and height in pixels three times per second. When a participant opened a course page in their browser and whenever they resized the browser window, the document size and the positions of all paragraphs in the text were logged. The maximum width of the text area was restricted to 900 pixels so that very wide screens would not display the full text on one page. The texts' paragraphs were annotated so that each paragraph had a unique identifier.

Only texts that were long enough so that students had to scroll to get from the start of the text to its end were included in the analysis. This was not the case for three of the texts in unit 3, which turned out to be too short. Furthermore, the tracking script was modified in the first week of the course, making the interactions with the five texts of the first unit unusable. Ultimately, 8 out of the 16 texts ended up in this analysis. As an indication of the text difficulty, we calculated the

Flesch Reading Ease score [21] for German texts with the *textstat* package. The score varied between 17.1 and 44.6, indicating that the texts were rather difficult (see Table 1).

2.3 Segmenting the Events into Sessions

Since each user could theoretically open a text as often as they wished, the data was segmented into text-sessions for each user and each text. A new text-session started whenever a user opened a different learning material. For example, a learner who opens the first document for 5 minutes, then the second document for 10 minutes, then the first document again for 10 minutes has three text-sessions: Two on the first document and one on the second document.

Text	# Words	# Sessions	Paragraph Ratio M (SD)	Flesch score
Unit2-1	978	507	.68 (.35)	31.7
Unit2-2	966	421	.75 (.37)	32.0
Unit2-3	949	429	.73 (.36)	27.9
Unit3-1	664	363	.78 (.35)	44.6
Unit3-2	649	341	.70 (.37)	34.2
Unit5-1	1485	415	.60 (.40)	30.0
Unit5-2	1340	442	.55 (.40)	25.3
Unit5-3	1578	342	.60 (.35)	17.1

Table 1. Descriptions of the texts that were part of the dataset.

We removed text-sessions shorter than 20 seconds. These text-sessions were very short and included only few interaction patterns with barely any scrolling, indicating that students did not make serious attempts to work with the texts.

2.4 Indicator of Disengagement

We created a paragraph-ratio indicator for each text-session to determine whether or not students were disengaged in a text-session. For constructing this indicator, we first used a uniform viewport attention model [11] to determine whether a paragraph was visible long enough to be thoroughly read. A uniform attention model assumes that a reader attends to all areas of the viewport with the same level of attention. Although there are other models that, for example, give the center of the viewport a higher weight [11], we wanted to keep the model as simple as possible.

Next, we computed the expected reading time for each paragraph as the word length divided by an estimate of average reading speed in words per minute (WPM). We used a value of 260 WPM based on the average reading speed identified for the silent reading of German texts across 13 studies [3].

Finally, the ratio of paragraphs that were visible long enough, divided by the number of paragraphs in the text, is the paragraph-ratio indicator for each text-session. If the paragraph-ratio-indicator is below 1, a text-session counts as not completed. The paragraph-ratio-indicator captures two types of behavioral disengagement: A. scrolling over a paragraph and B. abandoning the text by leaving it before reaching its end.

2.5 Time Series Representation

To transform the scrolling data into a time series representation, we first resampled the scrolling data with a frequency of 1Hz, so that one data point equals one second. For example, a series with 30 data points is also 30 seconds long. This has the advantage that the amount of data becomes smaller, which positively affects the method's time and memory requirements. Then we calculated the viewport document percentage at each sampling point by dividing the document height by the viewport top position.

2.6 Classification Procedure

Due to the rationale that early detection is essential for timely feedback, the classification of disengagement should happen as early as possible. Thus, we trained our classifiers with different time series lengths from the beginning of the respective session according to the following scheme: starting at the beginning of the session, we classified time series of increasing length, incremented by 10 seconds. Effectively, we classified the first 10 seconds, then the first 20

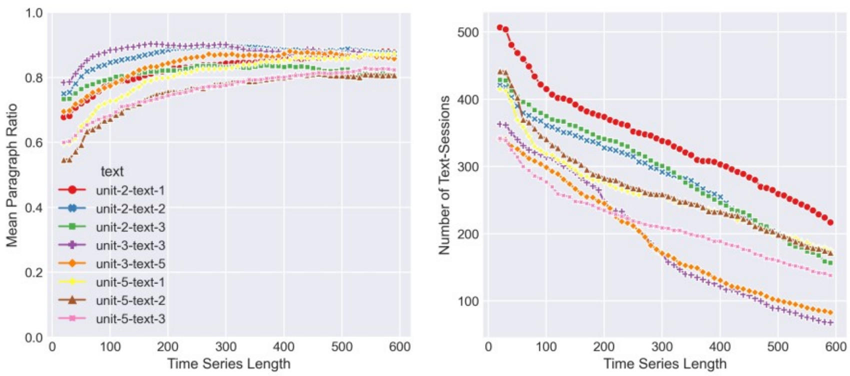


Fig. 2. Changes in the mean paragraph ratio (left) and the number of text-sessions (right) with increasing time series length.

seconds, all up to the first 150 seconds. We decided to use a maximum of 150 seconds because this was approximately the expected time to completely read the shortest text.

All programming was done in python. For the selection of algorithms, we used classifiers from the sktime package [15, 16], which offers state-of-the-art time series algorithms and is used in benchmarks of classifier performance [19].

2.7 Sampling and Participants

We aimed to classify time series of different lengths to assess how early the best possible classification is possible. However, due to different reading speeds and the fact that readers disengage from the texts at different times, there are sessions shorter than the time series used for some classifications. For example, if the algorithm is to classify time series of length 40s, then all those sessions shorter than 40s are dropped from the corresponding data set. Thus, the classifier trained on 30s time series uses a larger data set than the classifier trained on 40s time series.

Furthermore, the ratio of completed to uncompleted text sessions also increases, as the length of the text sessions increases. For the example above, we can assume that sessions shorter than 40s are not sufficient to completely read the text. Therefore, when we exclude shorter sessions from the classification, the class distribution in the data set changes.

The impact of time series length on data set size and class distribution is depicted in Figure 2. To investigate the consequences of these differences, we compared two sampling strategies: a dynamic sampling strategy, in which the data set consists of all eligible sessions for each time series length, and a static sampling strategy, where the data set contains only sessions that are 150s or longer so that classifications for different time series lengths all use the same data.

To address the uneven class distribution, we undersampled from the majority class such that there is always an even class distribution. With an even class distribution, there are less restrictions about the classification algorithms that can be used, as some do not work well on skewed class distributions. Moreover, the results are more straightforward to interpret, as we have a clear baseline of 50% random chance against which to compare the classifiers. To control the distortions due to undersampling, the classification runs ten times, each time independently sampling anew.

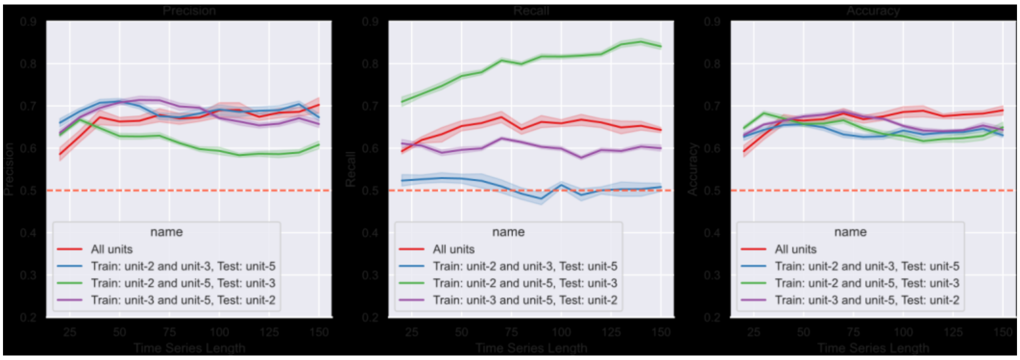


Fig. 3. Comparison of different classification algorithms, from left to right: Precision, Recall, and Accuracy. The MiniRocket and STSF algorithms achieve the highest precision, while the WEASEL algorithm has the highest recall on long time series, indicating fewer false negatives. The y-axes are truncated below 0.2 and above 0.8 for better visibility of the differences. The dashed red line indicates the result of randomly guessing. The shaded areas indicate the 95% confidence interval.

3 RESULTS

3.1 Comparison of classifiers

We evaluated classifiers from sktime version 0.13.0. Note that we report only about those that performed well or are otherwise noteworthy.

The best-performing algorithms turned out to be the kernel-based Minimally Random Convolutional Kernel Transform (MiniRocket) [7] and the interval-based Supervised Time Series Forest (STSF). In terms of accuracy, both performed within the margin of error of the other (see figure 3. STSF achieved higher precision (M = 0.68, SD = 0.04, Max = 0.72 at 150s, Min = 0.62 at 20s), and MiniRocket higher recall (M = 0.64, Max = 0.67 at 70, Min = 0.59 at 20). Only in recall of longer time series did the dictionary-based Word Extraction for Time Series Classification (WEASEL) algorithm perform better than the other algorithms (Max = 0.63 at 150s). In the combined accuracy metric, MiniRocket and STSF are barely distinguishable. We decided to use MiniRocket for further analyses because it had a significantly shorter total training time (609s) compared to STSF (4749s). Note that for this comparison, we used the static sampling strategy. The comparison of both strategies will give reason for this choice.

We compared the two sampling strategies, dynamic and static sampling (see section 2.7). With dynamic sampling of training data, the classification performance varies more than with static sampling, primarily in the recall metric (i.e., the classification results in more false negatives). This is illustrated in figure 4. Despite the dynamic sampling's better maximum precision, we adopted the static sampling strategy for further analyses to obtain a uniform sample and higher reliability across all-time series length.

To investigate the classification performance across different texts, we trained the classifier on the texts from two units while testing on a third. Once again, the accuracy was above chance for all configurations (see figure 5). The biggest difference was in the detection of false negatives (recall). The classification trained on unit-2 and unit-5 achieved higher recall on average (M = 0.79, SD = 0.04) and at maximum (Max = 0.85 at 140s) than the other configurations. Conversely, classifiers trained on unit-2 and unit-3, and tested on unit-5 had recall values around chance levels (M = 0.51, SD = 0.01, Max = 0.52 at 40s, Min = 0.48 at 90s).

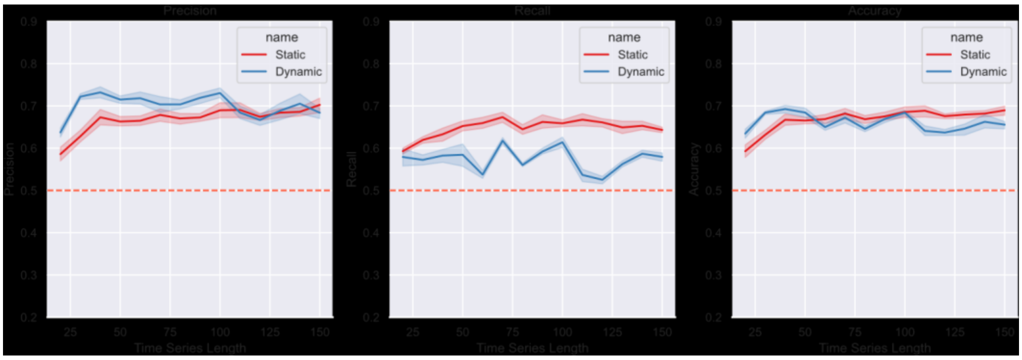


Fig. 4. Comparison of dynamic and static sampling of training data. Note that the y-axes are truncated below 0.2 and above 0.8 for better visibility of the differences. The dashed red line indicates the result of randomly guessing.

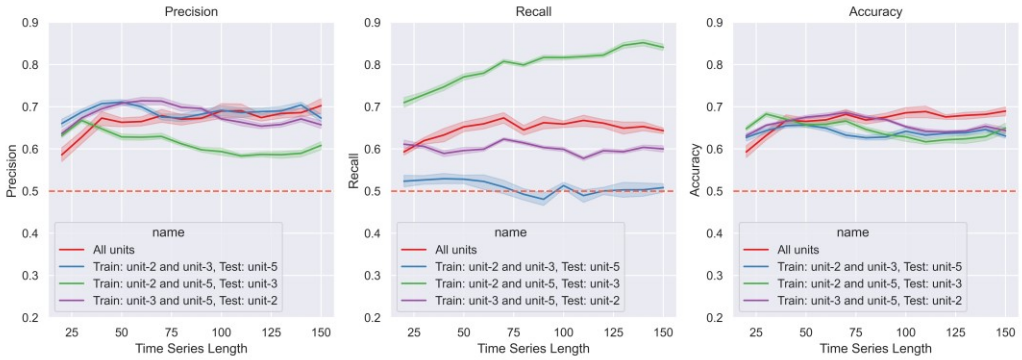


Fig. 5. Comparison of classification when training and testing on different texts. Note that the y-axes are truncated below 0.2 and above 0.9 for better visibility of the differences. The dashed red line indicates the result of randomly guessing.

4 DISCUSSIONS

We explored the early detection of disengagement of long texts in online learning environments using time series representations of scrolling data. With state-of-the-art classification methods, we achieved an accuracy of up to 70%. The classifier performance is comparable to what was achieved in previous works that also relied purely on trace data [18]. Our method has the advantage that it relies only on features that are readily obtained while students work with texts. Since this method had not previously been explored, we looked at the effects of the classification algorithm, time series length, and text complexity.

We found that the characteristics of the texts that are part of the training data have an impact on the classification performance. The configuration of training on texts from units 2 and 5 and testing on unit 3 stands out because it had the worst precision but by far the best recall performance. One reason could be that readers have to scroll more or less on different text lengths, thus effectively producing different scrolling patterns. Another explanation could be that the texts varied in difficulty, eventually producing different disengagement patterns.

4.1 Improvements and Recommendations for Future Work

The method presented in this study can detect disengagement behavior early in the reading process. Therefore, it yields a promising potential for early interventions to prevent readers from disengaging from a text. The peak accuracy of the classification was at 60s after the beginning of a text-session. Increasing the analyzed time series length beyond this did not result in further improvements. Nonetheless, further research is needed to extend our knowledge about and further improve the classification of disengagement behavior in reading tasks. We want to propose three promising avenues to achieve this: 1. Considering individual reading speed, 2. Exploring different attention distributions, 3. Improving the classification label of disengagement.

Regarding the individual reading speed, the tracking of paragraph completion, as we presented it, assumes the same reading speed for everyone, while individual reading speed can vary greatly [3]. We considered inferring an individualized reading speed based on the students' time on texts.

However, since we had no evidence of whether students actually invested the time in reading or in one of many other activities [4], we did not introduce this additional variability.

With respect to the attention levels, we assumed uniform attention over the whole screen for the inference of attention. It might be beneficial to evaluate different models of attention, for example, Gaussian distributions around the center of the screen [11].

Finally, we expect the most significant improvements to the method to come from better disengagement labels, where we relied entirely on trace data. For example, in our dataset, there are often long phases of inactivity (sometimes for several hours). We suspect that these phases of inactivity reflect a type of behavioral disengagement. With additional means of tracking, especially mobile devices, these phases of inactivity on the text could be mapped to the actual behavior outside the text and, therefore, lead to better class labels for the classification.

4.2 Limitations

Our participants and the texts were all part of the same course. All limitations to the generalizability of research in the social sciences apply. Different texts and different participants can be expected to lead to slightly different classification performances. Due to time and technological constraints on our side, we could not test convolutional neural networks for time series classification. Although reviews have shown that deep learning is no better than conventional approaches, evaluating their performance would still be interesting.

Despite these shortcomings and limitations, we see a lot of potential in our method for detecting disengagement behavior in reading tasks. The presented study has shown that our approach is applicable in natural settings of online learning, and with the consideration of possible future research, we expect that its performance can be further improved.

REFERENCES

[1] Anthony Bagnall, Jason Lines, Aaron Bostrom, James Large, and Eamonn Keogh. 2017. The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. *Data Mining and Knowledge Discovery* 31, 3 (May 2017), 606–660. https://doi.org/10.1007/ s10618-016-0483-9

[2] Daniel Biedermann, Jan Schneider, and Hendrik Drachsler. 2021. Digital self-control nterventions for distracting media multitasking - A systematic review. *Journal of Computer Assisted Learning* 37, 5 (Oct. 2021), 1217–1231. https://doi.org/10.1111/jcal.12581

[3] Marc Brysbaert. 2019. How many words do we read per minute? A review and meta-analysis of reading rate. *Journal of Memory and Language* 109 (Dec. 2019), 104047. https://doi.org/10.1016/j.jml.2019.104047

[4] Charles Calderwood, Phillip L. Ackerman, and Erin Marie Conklin. 2014. What else do college students “do” while studying? An investigation of multitasking. *Computers & Education* 75 (June 2014), 19–29. https://doi.org/10.1016/j.compedu.2014.02.004

[5] Lucy Chipchase, Megan Davidson, Felicity Blackstock, Ros Bye, Peter Colthier, Nerida Krupp, Wendy Dickson, Deborah Turner, and Mark Williams.

2017. Conceptualising and Measuring Student Disengagement in Higher Education: A Synthesis of the Literature. *International Journal of Higher Education* 6, 2 (March 2017), 31. https://doi.org/10.5430/ijhe.v6n2p31

[6] Mihaela Cocea and Stephan Weibelzahl. 2011. Disengagement Detection in Online Learning: Validation Studies and Perspectives. *IEEE Transactions on Learning Technologies* 4, 2 (April 2011), 114–124. https://doi.org/10.1109/TLT.2010.14

[7] Angus Dempster, Daniel F. Schmidt, and Geoffrey I. Webb. 2021. MINIROCKET: A Very Fast (Almost) Deterministic Transform for Time Series Classification. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 248–257. https://doi.org/10.1145/3447548. 3467231 arXiv:2012.08791 [cs, stat].

[8] Sidney K. D'Mello and Caitlin S. Mills. 2021. Mind wandering during reading: An interdisciplinary and integrative review of psychological, computing, and intervention research and theory. *Language and Linguistics Compass* 15, 4 (2021), e12412. https://doi.org/10.1111/lnc3.12412 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/lnc3.12412.

[9] Noah D. Forrin, Caitlin Mills, Sidney K. D'Mello, Evan F. Risko, Daniel Smilek, and Paul Seli. 2021. TL;DR: Longer Sections of Text Increase Rates of Unintentional Mind-Wandering. *The Journal of Experimental Education* 89, 2 (April 2021), 278–290. https://doi.org/10.1080/00220973.2020.1751578 Publisher: Routledge _eprint: https://doi.org/10.1080/00220973.2020.1751578.

[10] Patricia Goedecke, Daqi Dong, Genghu Shi, Shi Feng, Evan Risko, Andrew Olney, Sidney mello, and Arthur Graesser. 2015. *Breaking Off Engagement: Readers' Disengagement as a Function of Reader and Text Characteristics*.

[11] Max Grusky, Jeiran Jahani, Josh Schwartz, Dan Valente, Yoav Artzi, and Mor Naaman. 2017. Modeling Sub-Document Attention Using Viewport Time. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. Association for Computing Machinery, New York, NY, USA, 6475–6480. https://doi.org/10.1145/3025453.3025916

[12] Vitomir Kovanović, Dragan Gašević, Shane Dawson, Srećko Joksimović, Ryan S. Baker, and Marek Hatala. 2015. Penetrating the Black Box of Time-on-task Estimation. In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge (LAK '15)*. ACM, New York, NY, USA, 184–193. https://doi.org/10.1145/2723576.2723623 event-place: Poughkeepsie, New York.

[13] Dmitry Lagun, Chih-Hung Hsieh, Dale Webster, and Vidhya Navalpakkam. 2014. Towards better measurement of attention and satisfaction in mobile search. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*. ACM, Gold Coast Queensland Australia, 113–122. https://doi.org/10.1145/2600428.2609631

[14] Dmitry Lagun and Mounia Lalmas. 2016. Understanding User Attention and Engagement in Online News Reading. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*. ACM, San Francisco California USA, 113–122. https://doi.org/10.1145/2835776.2835833

[15] Markus Löning, Anthony Bagnall, Sajaysurya Ganesh, Viktor Kazakov, Jason Lines, and Franz J. Király. 2019. sktime: A Unified Interface for Machine Learning with Time Series. https://doi.org/10.48550/arXiv.1909.07872 arXiv:1909.07872 [cs, stat].

[16] Markus Löning, Franz Király, Tony Bagnall, Matthew Middlehurst, Sajaysurya Ganesh, George Oastler, Jason Lines, Martin Walter, ViktorKaz, Lukasz Mentel, chrisholder, RNKuhns, Leonidas Tsaprounis, Taiwo Owoseni, Patrick Rockenschaub, danbartl, jesellier, eenticott shell, Ciaran Gilbert, Guzal Bulatova, Lovkush, Mirae Parker, Kejsi Take, Patrick Schäfer, Stanislav Khrapov, Svea Marie Meyer, AidenRushbrooke, oleskiewicz, Yi-Xuan Xu, and Afzal Ansari. 2022. alan-turing-institute/sktime: v0.13.2. https://doi.org/10.5281/zenodo.7017832

[17] Caitlin Mills, Nigel Bosch, Art Graesser, and Sidney D'Mello. 2014. To Quit or Not to Quit: Predicting Future Behavioral Disengagement from Reading Patterns. In *Intelligent Tutoring Systems*, David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Alfred Kobsa, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Demetri Terzopoulos, Doug Tygar, Gerhard Weikum, Stefan Trausan-Matu, Kristy Elizabeth Boyer, Martha Crosby, and Kitty Panourgia (Eds.). Vol. 8474. Springer International Publishing, Cham, 19–28. https://doi.org/10.1007/978-3-319-07221-0_3 Series Title: Lecture Notes in Computer Science.

[18] Caitlin Mills and Sidney D'Mello. 2015. *Toward a Real-Time (Day) Dreamcatcher: Sensor-Free Detection of Mind Wandering during Online Reading*. International Educational Data Mining Society. https://eric.ed.gov/?id=ED560533 Publication Title: International Educational Data Mining Society.

[19] Alejandro Pasos Ruiz, Michael Flynn, James Large, Matthew Middlehurst, and Anthony Bagnall. 2021. The great multivariate time series classification bake off: a review and experimental evaluation of recent algorithmic advances. *Data Mining and Knowledge Discovery* 35, 2 (March 2021), 401–449. https://doi.org/10.1007/s10618-020-00727-3

[20] Lennard Schmidt and Erik Maier. 2020. *Assessing Ad Attention through Clustering Viewport Trajectories*.

[21] John Talburt. 1985. The Flesch index: An easily programmable readability analysis algorithm. In *Proceedings of the 4th annual international conference on Systems documentation - SIGDOC '85*. ACM Press, Ithaca, New York, United States, 114–122. https://doi.org/10.1145/10563.10583 [22] Maryanne Wolf and Mirit Barzillai. 2009. The importance of deep reading. *Educational Leadership* 66, 6 (March 2009).

Paper 3

Biedermann, D., Schneider, J., and Drachsler, H. (2021). *Digital Self-Control Interventions for Distracting Media Multitasking — A Systematic Review*. Journal of Computer Assisted Learning, 37(5), 1217–1231. https://doi.org/10.1111/jcal.12581

Digital self-control interventions for distracting media multitasking: A systematic review

Daniel Biedermann[1], Jan Schneider[1] and Hendrik Drachsler[1,2]
[1] Information Center for Education, DIPF | Leibniz Institute for Research and Information in Education, Frankfurt am Main, Germany
[2] Computer Science Faculty, Goethe University Frankfurt, Frankfurt am Main, Germany

Abstract

Digital distractions can interfere with goal attainment and lead to undesirable habits that are hard to get red rid of. Various digital self-control interventions promise support to alleviate the negative impact of digital distractions. These interventions use different approaches, such as the blocking of apps and websites, goal setting, or visualizations of device usage statistics. While many apps and browser extensions make use of these features, little is known about their effectiveness. This systematic review synthesizes the current research to provide insights into the effectiveness of the different kinds of interventions. From a search of the ‘ACM’, ‘Springer Link’, ‘Web ofScience’, ’IEEE Xplore’ and ‘Pubmed’ databases, we identified 28 digital self-control interventions. We categorized these interventions according to their features and their outcomes. The interventions showed varying degrees of effectiveness, and especially interventions that relied purely on increasing the participants' awareness were barely effective. For those interventions that sanctioned the use of distractions, the current literature indicates that the sanctions have to be sufficiently difficult to overcome, as they will otherwise be quickly dismissed. The overall confidence in the results is low, with small sample sizes, short study duration, and unclear study contexts. From these insights, we highlight research gaps and close with suggestions for future research.

1 Introduction

n learning scenarios such as blended learning or online courses, where learning is largely self-directed, digital content has a Janus-faced role. On the one hand are the potential benefits, such as the access to knowledge resources and the opportunity to exchange information quickly. On the other hand, there are many ways that learners can be distracted from their learning goals, for instance by watching entertaining videos, browsing social networks, or playing video games. Both the goal-congruent and the goal-incongruent content are often equally accessible. However, while learning can be unpleasant and exhausting, the distracting entertainment is engineered to be fun and addicting, to tempt the users to spend as much time on a platform as possible (Eyal, 2014).

Awareness of the detrimental effects of digital distractions is widespread. When students were asked about their smartphone usage, more than 60% responded that they felt they were overusing their smartphone and that it was a distraction in class (Ko et al., 2015). In an observation of learner behaviour, Rosen et al. (2013) observed that participants switched tasks on average every 6 min, mostly to social media. They also observed negative associations between the frequency of switching to social media and the GPA. This negative association between the frequency of using digital distractions during learning and academic performance was observed in several other studies (Kirschner & Karpinski, 2010; Masood et al., 2020).

Different disciplines use different terminology to describe the phenomenon of overuse of distracting content: The term 'media multitasking' is often used in the psychological research literature, and there are validated surveys to measure it (Baumgartner et al., 2017). While studies find overall that media multitasking is likely to be detrimental to academic performance (van der Schuur et al., 2015), there is still debate about positive effects of media multitasking. Another term that is often used to describe distractions at work is cyberloafing (Varol & Yıldırım, 2019). The term cyberloafing describes behaviour that is undesirable, especially from the perspective of third parties such as employers. Whether the behaviour referred to as cyberloafing is bad for the people themselves is, of course, another question. Meanwhile, in the human-computer interaction (HCI) literature, a broad range of terms like 'smartphone non-use' (Hiniker et al., 2016), 'self-interruption' (J. Kim et al., 2017), or 'self-control failure' (Lyngs et al., 2020) is used. Self-control has a clear construct and a model that further sheds light on the mechanisms behind frequent use of distracting content. Self-control intervention is also the term that is used for the digital interventions that are the subject of this review (Lyngs et al., 2019; Schwartz et al., 2021).

1.1 Theory and Related Works

If and how often someone follows a short-term gratifying impulse, instead of working towards their long-term goals, depends on several factors that are captured in the trait of self-control. Since this trait is relatively stable (Coyne & Wright, 2014) and even partially hereditary (Willems et al., 2019), some people are more susceptible to disrupting their learning than others.

Another factor that influences the outcome of self-control conflicts is habitual behaviour (Duckworth et al., 2019), which refers to behaviour that has reached a degree of automaticity such that it is no longer directed by conscious goals. It develops through frequent repetition of the same behaviour in a stable context (Fiorella, 2020; Mazar & Wood, 2018; Wood & Rünger, 2016). Future behaviour is then initiated by exposure to the same contextual cues, which can be locations, preceding actions, or mental states (Mazar & Wood, 2018; Wood & Rünger, 2016). This is especially challenging given the portable and ubiquitous nature of digital devices. Individuals can develop habitual behaviour in a wide range of contexts (Bayer & LaRose, 2018), and it is not unreasonable to expect that many disruptions of learning are due to the habitual use of digital devices. The often-habitual nature of using digital devices can also explain why especially 'heavy users' severely underestimate the amount of time they spend with digital distractions (H. Lee et al., 2014).

For both the conscious and the not-so-conscious uses of distractions, digital self-control interventions aim to provide support where the individual self-control is no longer sufficient. Lyngs et al. (2019) have conducted a review of consumer-oriented digital self-control interventions from the various app stores. In this review, the authors created an integrative model of self-control and habitual behaviour that aims to explain how these interventions can influence behaviour in the face of distractions (see Figure 1).

As a result of their work, Lyngs et al. (2019) identified 367 different interventions that they sorted into four broad categories, which reflect different targets where these interventions influence the decision to (not) use a distraction. Their categories are: (1) blocking/removing, (2) self-tracking, (3) goal advancement and (4) reward/punishment.

The interventions with features in the category blocking or removing block the access to programs or websites entirely, or they set limits on the duration that a user can access distracting content. In some cases, the user can pause the blocking with additional friction, like a menial task or a password. Other interventions in this category remove particularly distracting features, such as a newsfeed or content recommendations, from popular websites. These interventions can prevent habitual use of digital distractions, giving users the chance to reflect on their behavior and make a conscious decision instead of following an unconscious impulse.

If the behavior is driven by conscious decisions, the users can benefit from interventions with features from the self-tracking and goal advancement category. Self-tracking interventions display usage statistics or the amount of time that a user spent on distracting activities. Users are supposed to monitor themselves and to decide when they have overused digital content. Goal setting interventions remind the users of a task or a goal that they have set for themselves.

Lastly, the category reward/punishment contains interventions with features that provide rewards, points, or achievements that users can gain for abstaining from distractions. The rewards or punishments within these interventions are sometimes in the form of social actions in which users support or sanction each other. These features influence the value that users ascribe to resisting using a digital distraction (Shenhav et al., 2013).

While the review by Lyngs et al. (2019) resulted in an extensive overview of existing features, it is not known how effective they are at reducing the negative impact of digital distractions. In the context of self-directed learning with digital media, one can imagine scenarios for the various interventions in which they work well, or less well. For example, the approach of complete blocking does prevent distractions from being used. However, complete blocking also means that a learner can no longer access the recording of a course if it is hosted on the same video platform as the entertainment. For such scenarios, only interventions that still allow access to a platform are usable at all. These more permissive interventions then raise questions, like: How well do they prevent entertainment from being used instead of learning? Is it sufficient that learners can see their usage statistics in order to reduce distractions at critical moments? Are self-set time limit goals adhered to, or are additional restrictions necessary? With this review, we address these gaps by providing a systematic literature review of peer-reviewed publications that studied changes in the use of digital distractions when participants used digital self-control interventions.

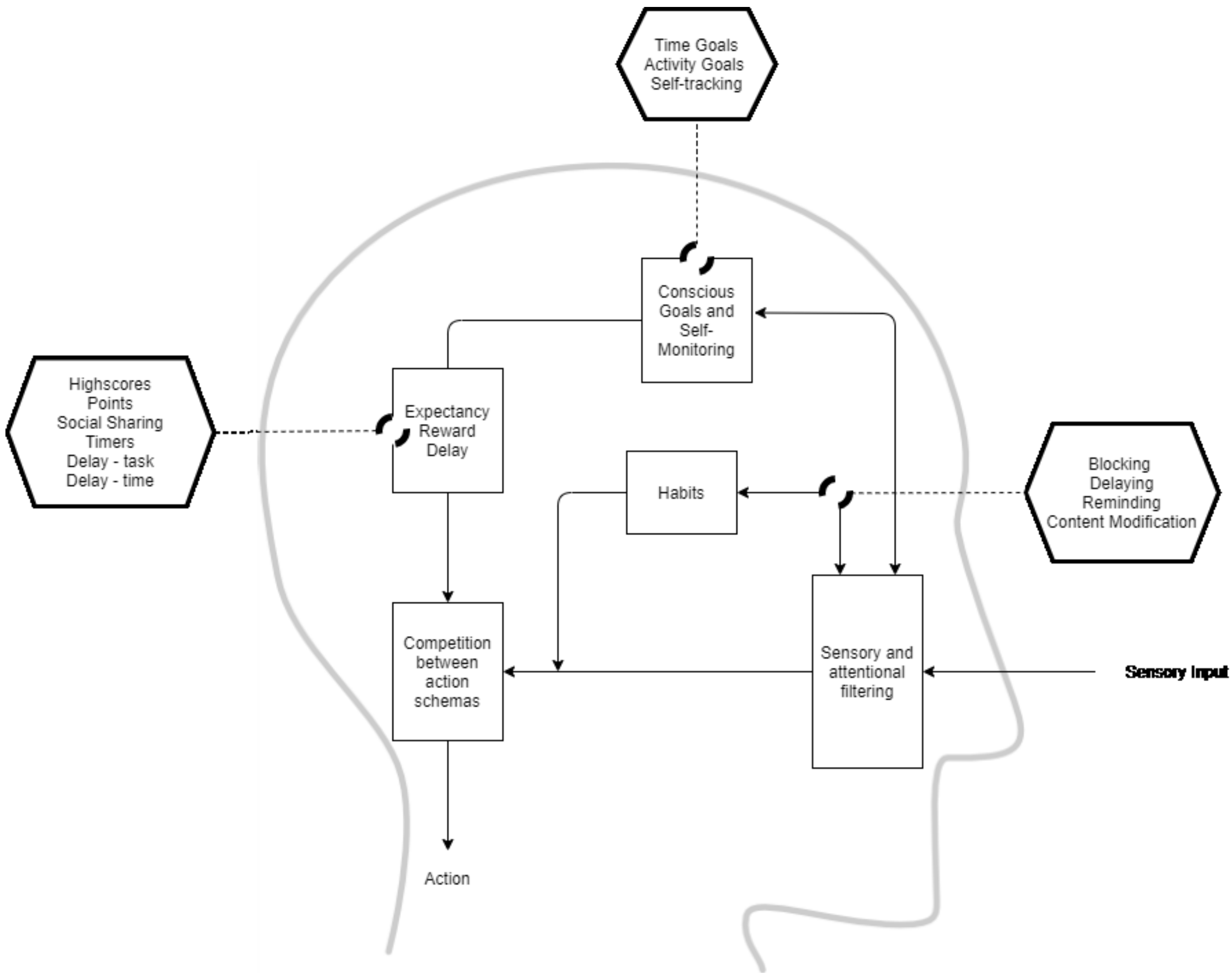


Figure 1: An extended dual systems model of self-control that Lyngs et al. (2019) adapted from Shea et al. (Shea et al., 2014). The hexagons list interventions that are connected via dashed lines to the targets where they intervene in the process from sensory input to action.

1.2 The Present Study

In this study, we contribute to the research on digital self-control interventions with a systematic review (Petticrew & Roberts, 2006) where we examine the effectiveness of these interventions and their underlying features. We used the following research questions to guide our study:

RQ 1: Which types of intervention have evidence for achieving changes in the use of digital distractions?

RQ 2: How effective were the interventions at alleviating digital distractions?

To answer RQ 1, we first looked for published research regarding digital interventions for digital distractions. We then categorized the identified interventions based on the categories proposed by Lyngs et al. (2019). We present these results in a narrative form (compare Petticrew & Roberts, 2006, p. 165f.) where we describe the features of the various interventions. To answer RQ 2, we analyzed the reported results and, when possible, calculated the effect sizes of the interventions (See the Data Extraction and Analysis section).

2 Method

For this review, we followed the Preferred Reporting Items for Systematic Reviews and Meta-Analysis guidelines (PRISMA) (Moher et al., 2009), which sets out the minimum standards of reporting that should be present in a systematic review, and a flow of information to apply for the study selection.

Eligibility criteria

We included publications that reported the effects of digital interventions to alleviate digital distractions. Our inclusion criteria were publications that…

…were written in English.

…were published in a peer-reviewed outlet.

…applied a digital intervention to reduce the use of distractions on a computer, smartphone, or a wearable device.

…reported outcome measures related to changes in the use of digital distraction. Since an effect for a reduction in use is guaranteed for interventions that completely block the distractions, we included use of the intervention as an outcome. The rationale is that if the blocking is used more often, distractions can be accessed less frequently.

…had a study design that compared outcomes under treatment with the outcomes of a non-treatment group, or a single-group design with a baseline measurement.

…investigated a nonclinical population, for example, no special needs or ADHD population.

2.1 Search Strategy

We performed a search of the relevant databases and combined these results with a snowballing approach to obtain additional literature from the references and the citing literature (see Figure 2). For the database search, the search concepts were (1) digital distractions and (2) interventions. From the terminologies used in the literature, we arrived at the search term (‘self-control’ OR ‘self-regulation’ OR ‘mind-wandering’ OR cyberloafing OR ‘media multitasking’ OR willpower OR ‘digital distractions’ OR ‘smartphone addiction’) for the topic of digital distractions. For the interventions, the terms were (intervention* OR app OR browser* OR tracker OR self-monitoring OR digital

OR smartphone). Both terms were combined with a logical AND. We searched the databases with the broadest applicable search scope. The databases were: ‘Web of Science’ (scope: Topic), ‘Springer Link’ (scope: All), ‘ACM Digital Library’ (scope: Anywhere), ‘IEEE XPlore’ (scope: Full Text and Metadata) and ‘Pubmed’ (scope: All Fields). We copied all the results into a spreadsheet, where the publications were filtered first by title. Next, we read the abstracts and, for those publications which appeared eligible based on their abstract, we read the full text. From the results of the database search that we considered as relevant, we looked for further papers by incorporating the referenced literature and the literature which cited the publication. For this step, we used ‘Google Scholar’, which had the most extensive results for citing literature.

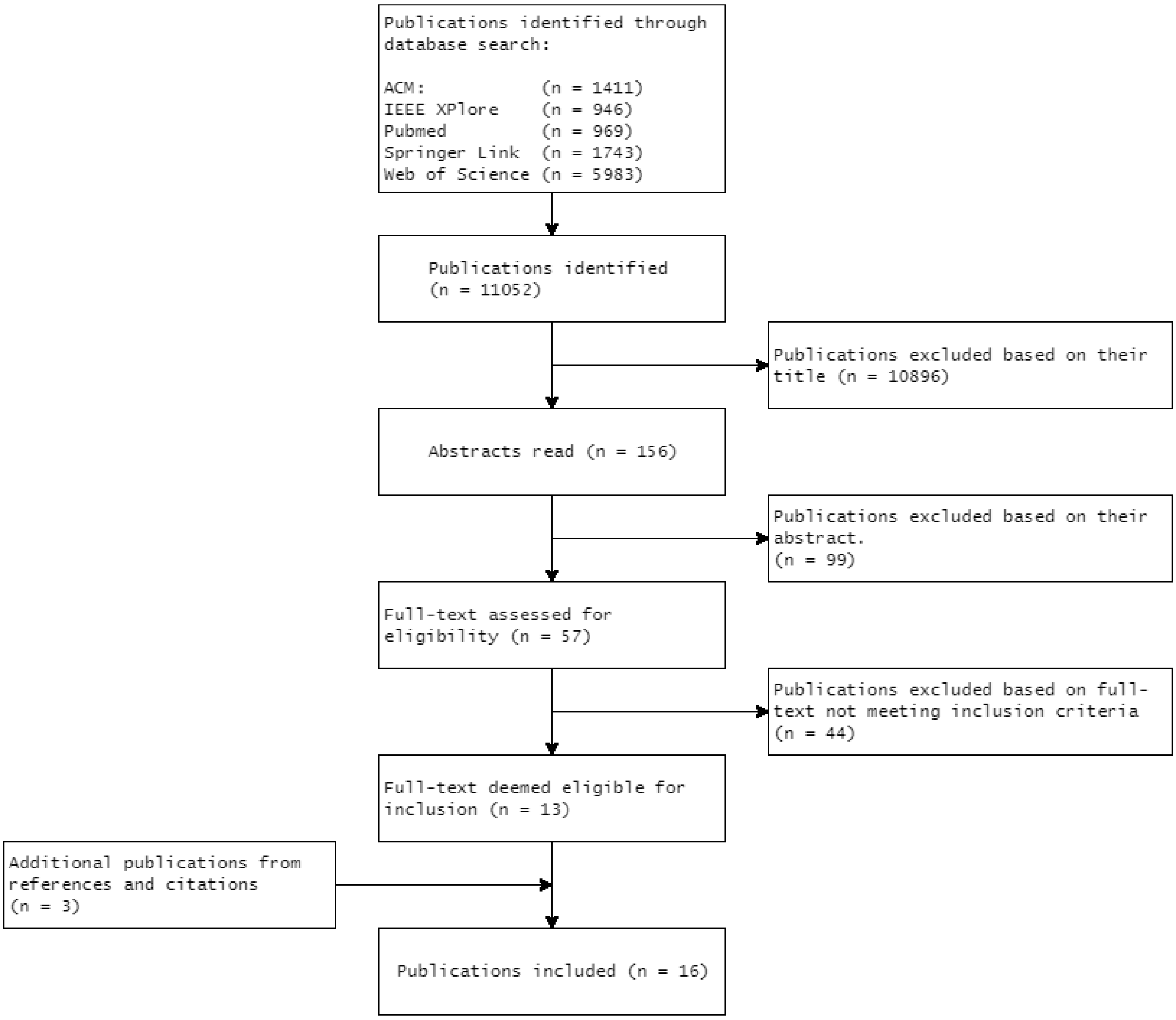


Figure 2: Flowchart of the paper inclusion process

3 Data extraction and analysis

For each publication, we extracted a description of the intervention(s), the research questions or hypotheses, the outcomes, the year of publication, the study design, the study population, recruitment criteria, the study duration, the target distraction, and the measured effects of the intervention(s). If a publication contained multiple interventions or multiple studies for one intervention, we extracted the parameters mentioned above separately for each intervention and each study.

To categorize the features, the first and second authors of the paper independently rated the interventions in the publications according to the coding criteria. The inter-rater agreement between authors was 91%. Consensus about the remaining 9% of the categorizations was reached after a brief discussion.

To calculate effect sizes and confidence intervals, we used the following versions of the calculation of standardized mean difference Cohen's d. For between subjects-design we calculated the difference of the means (M) divided by the average standard deviation (SD) (Lakens, 2013):

$$\frac{M_1 - M_2}{(\frac{SD_1 + SD_2}{2})}$$

For within-subjects design, we calculated the effect size as the difference of the means divided by the pooled SD (Lakens, 2013):

$$\frac{M_1 - M_2}{(\frac{\sqrt{(n_1 - 1)SD_1^2 + (n_2 - 1)SD_2^2 +}}{n_1 + n_2 - 2})}$$

For publications that did not allow conversion to these effect size measures because of missing data in their report, we tried to obtain the data by contacting the corresponding authors of the publication.

3.1 Risk of Bias and Quality Assessment

To assess the study quality and the risk of bias in the individual studies, we followed the study quality assessment tools by the NHLBI[8]. In addition to the criteria from those checklists, we considered study duration, the conduct of a follow-up, and comprehensible reporting of results. This assessment only determines how well the studies are suited to answer our research questions and does not represent a general categorization into "good" or "bad" studies.

3.2 Search Results

We found 16 publications eligible for this review, which contained a total of 28 interventions. The interventions are listed with a brief description in Table 1. We assigned an identifier consisting of the number sign # and a number to each intervention (#1–#16). Nine publications contained more than one intervention and we appended a lower-case alphabetic letter to their id (e.g., #7a, #7b).

Table 1

The publications that were included in this review and a description of the respective interventions.

ID	Reference	Description
#1	(Lottridge et al., 2012)	Browser redesign to frame tabs either as "work" or "non-work". "Work" tabs were larger, while other tabs were smaller and placed on the right side. A "work vs non-work" ratio progress bar was added to the browser status bar.
#2a #2b	(Ko et al., 2015)	*NUGU* - A smartphone app that encouraged participants to set time goals during which they commit to not use their smartphone. Comparison of a social comparison feature (#2a) with a variant of the app without this feature (#2b).
#3a #3b	(Foulonneau et al., 2016)	*TILT* - A smartphone app that displayed device use in several visualizations: messages, notifications and home screen widgets. Content was personalized (#3a) or static (#3b).

[8] https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools

#4	(Hiniker et al., 2016)	*MyTime* - An app to set time limit goals for monitored apps. Upon reaching those goals, reminders to stop using those apps popped up.
#5a #5b	(Y.-H. Kim et al., 2016)	*TimeAware* - Information display of time spent on the device with either a productivity (#5a) or a distraction (#5b) framing.
#6	(Terry et al., 2016)	Daily text messages with facts about the negative effects of media multitasking.
#7a #7b	(Whittaker et al., 2016)	*meTime* - An always-on dashboard that shows the last used applications. Longer use of an application resulted in a larger display of this application in the dashboard.
#8	(J. Kim et al., 2017)	*PomodoLock* - Distracting apps and websites were blocked for 25-minute intervals on both smartphone and PC.
#9a #9b	(Kovacs et al., 2018)	*HabitLab* - A browser extension with many different interventions. Study #9a investigated the effect of rotating interventions on attrition, and study #9b the effect of providing additional information and control over the interventions.
#10a #10b	(Okeke et al., 2018)	Repeating smartphone vibrations when a monitored website was visited longer than allowed (five minutes #10a, personalized #10b).
#11a #11b #11c	(J. Kim, Park, et al., 2019)	*Lock n' Type* - Participants had to enter a string of random digits before opening monitored apps. Digit input lengths were zero (#11a), ten (#11b), and 30 (#11c).
#12a #12b #12c	(J. Kim, Jung, et al., 2019)	Goalkeeper - different lockout mechanisms based on individual time goals. From reminder notifications (#12a), to successively longer lockout durations (#12b), and full-day phone lock upon exceeding the time (#12c).
#13a #13b	(Tseng et al., 2019)	*UpTime* - Conversational agent-based website blocker (#13a) compared with a time-based website blocker

		(#13b). With #13a, users had to provide reasons for visiting distracting websites.
#14	(Holte & Ferraro, 2020)	Changing the display colors of Smartphones to grayscale to decrease the attractiveness of distracting content.
#15	(Loid et al., 2020)	Evaluation of the effect of push notifications on screen time and phone checking behavior.
#16a #16b	(Lyngs et al., 2020)	Two alterations of the *Facebook* website. In #16a, users had to set action goals when entering the site; In #16b, the newsfeed was removed.

4 Findings and Analysis

We start this section with a high-level overview of the interventions and the outcomes that the studies report, followed in the subsections by a narrative description of the interventions. Regarding the multiple interventions within one publication, it should be noted that some interventions within one publication often differed only in minor details, such as the time spent on a distraction before an intervention started. With a few exceptions, we categorized the interventions following the taxonomy that was identified in the review of the commercially available interventions (Lyngs et al., 2019) that we describe above. We deviated in the two following categories: First, we use the term 'awareness' (cf. Parry & le Roux, 2019) instead of 'self-tracking' because the concept of increasing the users' awareness is more appropriate. Second, we placed interventions that modified the features or content into their own category instead of together with the blocking features. Blocking is functionally different from the modification of features because blocking completely prevents access to the distraction. Feature modification comes into effect only after the content has already been accessed, and a change in activities has already taken place.

The interventions monitored and targeted different kinds of digital activities on different devices. The monitored activities were either smartphone apps, computer programs, or websites (see Table 3). The monitoring was mostly limited to the start of the activity and the duration with which it was active. In one publication, the analysis was extended to more fine-grained interactions within the distraction#16a,#16b. The interventions were active on either mobile devices (primarily smartphones) and computers. Only one intervention synchronized the monitoring across smartphones and computers. From the interventions on mobile devices, six of the interventions monitored all app activity, nine monitored activity only for apps from a user- or application-defined list, and two interventions monitored only the activity in one specific app. All of these interventions that ran on a computer were designed to monitor website activity, and four interventions also monitored the programs that the participants used. Eleven of the interventions which monitored the browser activity used specific user- or application-defined URLs as intervention targets, while two interventions monitored only the activity on a specific website. The following subsections give more details regarding the description of the intervention features and the outcomes that were achieved (Figure 3).

Figure 3

Parameters that describe the various interventions.

			Feature Categories					Outcome Categories			
ID	**Device**	**Monitored Activities**	**Aware**	**Block**	**Goals**	**Rewards & Social**	**Content Mod.**	**Time on distraction**	**Distraction start**	**Time on device**	**Intervention use**
1	💻	Specific websites	■				■	+	+		
2a	📱	All apps	■		■	■			+	+	+
2b	📱	All apps	■		■	■			+	n.s.	n.s.
3a	📱	All apps	■							n.s.	
3b	📱	All apps	■							n.s.	
4	📱	Specific apps			■			n.s.			
5a	💻	All websites, programs	■					n.s.			n.s.
5b	💻	All websites, programs	■					+			n.s.
6	📱	Specific apps	■					+			
7a	💻	All websites, programs	■					+		+	
7b	💻	All websites, programs	■					+		+	
8	💻 / 📱	Specific websites, apps		■	■						n.s.
9a	💻	Specific websites	■				■	+			-
9b	💻	Specific websites	■				■				+
10a	📱	Facebook app	■					+			
10b	📱	Facebook app	■					+			
11a	📱	Specific apps		■				-	+	n.s.	
11b	📱	Specific apps		■				-	+	n.s.	
11c	📱	Specific apps		■				-	+	n.s.	
12a	📱	Specific apps	■		■					n.s.	
12b	📱	Specific apps		■	■					+	
12c	📱	Specific apps		■	■					+	
13a	💻	Specific websites		■				+	+		+
13b	💻	Specific websites		■	■			n.s.	n.s.		-
14	📱	All apps, websites					■	+		+	
15	📱	All apps, websites	■						n.s.	n.s.	
16a	💻	Facebook website			■			+	+		
16b	💻	Facebook website					■	n.s.	+		

Note. The 'Device' column lists the device type on which the intervention was active (computer or mobile device). The columns on monitored activities and feature categories list these parameters for each intervention. The columns under the 'Outcome Categories' header list the outcomes that were achieved with the intervention. A '+' signifies that the intervention had significant positive results on this outcome, a 'n.s.' that there were no significant effects on this outcome, and a '−' stands for negative effects on this outcome

4.1 Description of Intervention Features

4.1.1 Awareness Features

Interventions with awareness features inform the participants about their use of distracting activities. This information is presented with the aim to start a process of self-reflection so that the participants themselves decide to spend less time on digital distractions.

Awareness interventions varied according to (a) the mode in which they presented the information to the participants, whether it was within a program or an app, or via instant message, or as a push notification. (b) the content that was presented to the participants, whether it was statistics, messages, or graphics.

The most common form of presentation was the display of usage statistics: Whittaker et al. (2016) created a computer program[#7a,#7b] which displayed the most-used programs from the last 30 min to the participants, showing the most-frequently used programs larger than the less frequently used ones. Lottridge et al. (2012) modified the web browser to show statistics about the frequency of work versus non-work websites that the participants visited[#1]. Other interventions contained visualizations of usage statistics on top of various other features[#2a,#2b,#12a,#12b,#12c] that are described in the following sections.

Another common form of presentation were different kinds of notifications that alerted the participants: text messages[#6], push notifications[#3a,#3b,#15a], and popup message dialogues[#11a,#4,#12a]. The most insistent awareness intervention, created by Okeke et al. (2018), started smartphone vibrations when a time quota for a distracting website (in their case, Facebook) was exceeded[#10a,#10b]. The smartphone vibrated until the user left the distracting website.

With two exceptions, the interventions informed the participants about the extent of distracting activity use. The first exception were the SMS notifications in the intervention from Terry et al. (2016), which contained facts about the negative effects of media multitasking[#6]. The second deviation concerned the framing of the content, where all but one intervention focused on the distracting activities. To investigate the effect of a positive framing, Y.-H. Kim et al. (2016) used visualizations for the use of productive programs like word processors[#5a] and compared this with distraction-focused visualizations[#5b].

4.1.2 Goal-Advancement Features

We considered a feature as goal-advancement if participants had the option to set goals related to a reduction of time spent on digital distractions. The goals were either time-based or action-based, where the participants could enter their goals in a free text. The majority of the interventions with goal-advancement features asked the participant to set a time limit for their distracting activities[#2a,#2b,#4,#12a,#12b,#12c], and free text goals were less common[#16a,#4]. To sanction goal deviations, the interventions used blocking of distractions[#12b,#12c], repeated reminders[#16a,#12a,#4], or the reduction of a virtual score[#2a,#2b].

Kim, Jung, et al. (2019) investigated sanctions of exceeding time-goals with three different interaction lockout mechanisms (ILM): In the Non-ILM condition[#12a], users received only notifications that reminded them of their time limit. In the Weak-ILM condition[#12b], the participants' phones were locked for increasing durations, and Strong-ILM[#12c] locked the users' phones for the rest of the day after exceeding their time limit goal.

Hiniker et al. (2016) evaluated a smartphone app which supported goals that were both time- and action-based[#4]. Participants were asked to enter a time limit goal and an optional free text goal for the use of monitored apps. After exceeding the self-set time limit, the participants were reminded of the excess with a warning prompt. Following this prompt, they could close the app, request a time extension, or simply dismiss it.

Lyngs et al. (2020) modified the Facebook homepage such that participants were prompted to enter free text action goals on each visit#16a. The goal that the participants entered was periodically repeated during their visit on the website.

4.1.3 Blocking Features

An intervention was considered as containing blocking features if it prevented access to the distraction. For blocking interventions, we found interventions where the blocking was rigorous[#8,#10,#12b,#12c,#13b] and others where lifting of the block could be negotiated[#11b,#11c,#13a,#12b].

Mark et al. (2018) blocked distractions during work time with the 'freedom' software, a blacklist-based website blocker[#10]. Other interventions were less rigorous and blocked distractions only for specific time intervals. Kim, Cho et al. (2017) created a Pomodoro-timer intervention on the usage of distracting websites. In the Pomodoro method, users set a time interval (typically 25 min), during which they plan to work productively. After this time interval, they get a break, typically 5 min, during which they are free to do whatever they like, before starting the next Pomodoro session.

In two publications, the block was not absolute, and could be lifted: Tseng et al. (2019) created the UpTime-system[#13a], a website-blocking system with a chatbot to control a time-based blocking system for the transition from break to work. A break was operationalized as an absence of user input for at least 5 min. Upon resumption of work, the chatbot was activated and started disabling access to distracting websites for the next 25 min. The participants could negotiate a lift of the block by navigating through a conversation with the chatbot. The chatbot also suggested to start blocking sessions when a participant spent more than 15 min on distracting websites. This system was compared with a time-based blocking system[#13b] (cf. #8). Kim, Park, et al. (2019) created the Lock n' Type intervention, which required participants to enter a string of random digits on their smartphone before they could open a blacklisted app. They tested three different variants of this intervention, which each required different lengths of random digits[#11a,#11b,#11c].

4.1.4 Modification of Content

For this review, features were considered as 'Modification of Features or Content', if they modified certain aspects of a digital distraction to make it still usable, but less appealing to the participants. To achieve this, these interventions remove non-essential parts of the distraction that are typically designed to convince the users to spend more time with an app or a website.

Holte and Ferraro (2020) switched their participants' smartphone displays to grayscale[#14] in order to reduce gratification from distracting activities. The grayscale filter was an integrated feature of their participants' smartphones, and the filter was active at all times and for all apps. Lottridge et al. (2012) modified the web browser tab bar[#1] so that tabs that were classified as work-related had their colour contrast enhanced, and were made larger. Tabs from non-work URLs were made smaller and always displayed on the right.

Kovacs et al. (2018) report about "Habitlab", a browser extension that rotates behaviour change interventions for distracting websites, that is, the participants experienced different interventions between visits to the same site[#9a,#9b]. The authors are not specific about all of the interventions that were active in Habitlab at the time of their studies, but they do mention a few example interventions, like a news feed blocker for Facebook, or hiding ’related videos' on YouTube.

Lastly, Lyngs et al. (2020) used a browser extension to remove the newsfeed[#16b] from the Facebook homepage.

4.1.5 Reward and punishment

Ko et al. (2015) added social support and point reward features to their ’NUGU’ smartphone app. The participants were asked to specify a duration for which they wanted to stop using their smartphone (e.g., ‘studying for 30 minutes’). These limiting sessions were conceptualized as ‘missions’, and completing a mission gave the users virtual points. If a participant used the smartphone for anything except incoming calls, the mission failed. Participants could also form groups, start missions together, and compare their time-limiting efforts with those of their peers[#2a].

4.2 Outcomes

For the outcome measurements, we clustered the outcomes into the categories (a) time that participants spent on the distraction, (b) the frequency with which a participants started a distraction, (c) the total time that participants spent on the device, and (d) measures of using and interacting with the intervention.

Measures for the time that participants spend with the distraction are a straightforward way to gauge the effectiveness of an intervention. In this category, we included outcomes that measured how much time the participants spent on websites or in apps that were marked as a distraction by the participants themselves or by the researchers. Summarized in Table 2, the time spent on distraction was measured for 17 interventions.

Table 2

The measures for the category “time that participants spent on the distraction”

ID	**Outcome**	**Measure**	**Effect Size** [95% CI]
#1	Time on distraction	Time in non-work URLs	1.01 [0.22, 1.80]
#4	Time on distraction	Time in monitored apps	0.66 [0.07, 1.25]
#5a	Time on distraction	Time in monitored apps/URLs	n.s.

#5b	Time on distraction	Time in monitored apps/URLs	(+)
#6	Time on distraction	MTUAS Scale	n.s.
#7a	Time on distraction	Time on Facebook	0.85 [0.47, 1.22]
#7a	Time on distraction	Browsing time	0.74 [1.11, 0.37]
#7b	Time on distraction	Time on Facebook	(+)
#9a	Time on distraction	Time in monitored URLs	(+)
#10a	Time on distraction	Time on Facebook	0.49 [-0.23, 1.22]
#10b	Time on distraction	Time on Facebook	0.60 [-0.05, 1.25]
#11a	Time on distraction	Time in monitored apps	(-)
#11b	Time on distraction	Time in monitored apps	(-)
#11c	Time on distraction	Time in monitored apps	(-)
#13a	Time on distraction	Time in monitored URLs	0.14 [-1.04, 1.33]
#13b	Time on distraction	Time in monitored URLs	n.s.
#14	Time on distraction	Use of social media apps	0.44 [0.11, 0.77]
#14	Time on distraction	Use of video player	n.s.
#14	Time on distraction	Use of web browser	0.46 [0.13, 0.78]
#15	Time on distraction	Self-report	n.s.
#16a	Time on distraction	Time on Facebook	1.62 [0.90, 2.33]
#16a	Time on distraction	Facebook visit length	n.s.
#16b	Time on distraction	Time on Facebook	n.s.
#16b	Time on distraction	Facebook visit length	0.75 [0.09, 1.40]

Note. When the reporting did not allow the conversion into Cohen's d, the measure is marked with a (+) if a significant reduction in the time that participants spent on the distracting activity was reported. No significant effects for this measure are marked with a "n.s.", and instances where an increase in the time spent on the distracting activity was observed are marked with a (-).

Measures of the frequency with which distractions are started subsume measures such as navigating to a URL, starting an app, or unlocking the phone. Changes in the frequency of distraction starts were measured for nine interventions, summarized in Table 3.

Table 3.

The measures that we subsumed under the category frequency of distraction starts

ID	**Outcome**	**Measure**	**Effect Size** [95% CI]
#1	Distraction start	Navigation to monitored URLs	1.84 [0.56, 2.22]
#2a	Distraction start	Opening apps	0.56 [0.09, 1.04]
#2b	Distraction start	Opening apps	n.s.
#11a	Distraction start	Opening monitored apps	(+)
#11a	Distraction start	Discouraged app starts	(+)
#11b	Distraction start	Opening monitored apps	(+)
#11b	Distraction start	Discouraged app starts	(+)
#11c	Distraction start	Opening monitored apps	(+)
#11c	Distraction start	Discouraged app starts	(+)
#13a	Distraction start	Periods with visits to monitored URLs	0.64 [-0.56, 1.84]
#13b	Distraction start	Periods with visits to monitored URLs	n.s.
#15	Distraction start	Number of phone checks	n.s.
#16a	Distraction start	Number of visits to Facebook	1.62 [0.90, 2.33]
#16b	Distraction start	Number of visits on Facebook	n.s.

Note. When the reporting did not allow the conversion into Cohen's d but still reported significant reductions, the entry is marked with a (+). No significant effects are marked with a "n.s."

Time spent on device measures does not discriminate between different activities on the device and can also include the use of apps or programs related to work or learning. These measures include the time spent across all apps or on all URLs. Time spent on the device was measured for 16 interventions. The results are summarized in Table 4.

Table 4

The measures that we subsumed under the category time that participants spent on the monitored device

ID	Outcome	Measure	**Effect Size** [95% CI]
#2a	Time on device	Time in all apps	0.83 [0.34, 1.32]
#2b	Time on device	Time in all apps	n.s.
#3a	Time on device	Time in all apps	n.s.
#3b	Time on device	Time in all apps	n.s.
#4	Time on device	Average daily smartphone use	0.48 [-0.11, 1.07]
#7a	Time on device	Total time on PC	0.61 [0.24, 0.97]
#7b	Time on device	Total time on PC	(+)
#11a	Time on device	Total time on smartphone	n.s.
#11b	Time on device	Total time on smartphone	n.s.
#11c	Time on device	Total time on smartphone	n.s.
#12a	Time on device	Smartphone use on workdays	n.s.
#12a	Time on device	Smartphone use on weekends	0.56 [0.13, 0.99]
#12b	Time on device	Smartphone use on workdays	0.35 [-0.12, 0.82]
#12b	Time on device	Smartphone use on weekends	0.51 [0.09, 0.93]
#12c	Time on device	Smartphone use on workdays	0.54 [0.07, 1.01]
#12c	Time on device	Smartphone use on weekends	0.51 [0.10, 0.96]

#14	Time on device	Total time on smartphone	0.39 [0.06, 0.72]
#15	Time on device	Total time on smartphone	n.s.

Note. When the reporting did not allow the conversion into Cohen's d, the measure is marked with a (+) if a significant reduction of the time that the participants spent on the device was reported. No significant effects for this measure are marked with a 'n.s.'.

Measures of how a participant interacts with an intervention can also provide indications of the effectiveness of an intervention. For blocking interventions, an active intervention means that no distracting activity can be started in the first place. Thus, a participant that activated such an intervention more often, might have made the intervention more effective than a participant that rarely activated the intervention. These measures can, furthermore, hint as to whether the intervention failed because it was generally not suited to achieve the task, or because the participants simply did not use it. The results for this outcome category are summarized in Table 5.

Table 5

The measures subsumed under the category use of the intervention

ID	**Outcome**	**Measure**	**Effect Size** [95% CI]
#2a	Intervention use	Number of goals	(+)
#2b	Intervention use	Number of goals	n.s.
#5a	Intervention use	Dashbaord access	n.s.
#5a	Intervention use	Widget access	n.s.
#5b	Intervention use	Dashbaord access	n.s.
#5b	Intervention use	Widget access	n.s.
#8	Intervention use	Number of started sessions	n.s.
#9a	Intervention use	Installation survival	(-)
#9b	Intervention use	Installation survival	(+)
#13a	Intervention use	Number of sessions	2.06 [1.17, 2.94]

#13b	Intervention use	Number of sessions	n.s.

Note. When the reporting did not allow the conversion into Cohen's d, the measure is marked with a (+) if a significant reduction of the time that the participants spent on the device was reported. No significant effects for this measure are marked with a 'n.s.'.

4.2.1 Awareness Intervention Effects

Awareness interventions led to reductions in time spent on distractions and time on the device both through the display of usage statistics[#5a,#5b,#7a,#7b] and notifications[#10a,#10b].

The interventions #7a and #7b achieved a medium to large effect for time on distraction and total time on the device, but both interventions were active for only 2 days (see also Table 6). Dashboard interventions led to less time on distraction only with distraction-framed content[#5b], but not with a productive framing[#5a]. Neither distraction nor productivity framing[#5a,#5b] made a difference regarding how long and how often the participants accessed the information.

Awareness notifications only achieved positive effects when they were insistent. For usage monitoring via notifications, no effect was observed on total time on device[#3a,#15], and frequency of distraction starts[#15]. Regular messages about the negative effects of excessive media consumption also had no positive effects[#6]. Only the smartphone vibrations to remind of overuse led to less time spent on the distraction with personalized[#10a] (d = 0.49) and static time quotas[#10b] (d = 0.69).

4.2.2 Goal-Advancement Intervention Effects

Time-based goal setting with warning prompts led to reductions in time on device, and the amount of time spent on distractions in one intervention[#4], while in other interventions, these prompts of exceeding a time limit did not result in a reduction in time on device until the excess was sanctioned[#12a]. When sanctions in the form of device lockout[#12b,#12c] were added, time on device was reduced (#12b: d = 0.35, #12c: d = 54).

When participants set action goals in addition to time goals, they were more likely to follow prompts to leave a distraction[#4]. Setting action goals immediately before the start of a distracting activity#16a also led to a large reduction in total time spent on the distraction (d = 1.62) and the frequency of distraction starts (d = 1.62). Changes in the duration of individual sessions were not significant.

4.2.3 Blocking Intervention Effects

Blocking the access to distractions with a task convinced participants to not start distractions proportionally to the task difficulty: Requiring a 30-digit input before opening an app[#11c] was more effective (d = 0.47) than a 10-digit input task (d = 0.27) or a simple warning prompt (d = 0.13). This task-based delay was the only type of intervention for which an increase in time on distraction was reported, indicating that participants compensated for the initial hurdle of opening an app by spending more time in it. Still, the aspect of being able to negotiate the lift of a block was preferred[#13a] to time-based blocking, and participants activated the intervention more often (d = 2.06) and started distraction less often (d = 0.82) than with purely time-based blocking intervention[#13b]. However, the chatbot also suggested the initiation of additional blocking sessions, thereby potentially skewing the results.

A measure that is particularly relevant for blocking interventions is how frequent the blocking is activated. Participants in a blocking condition#8 completed more limiting sessions in the first week than a control group which only monitored themselves. Over the full two-week duration, these differences were no longer observable.

4.2.4 Content Modification Effects

Setting the smartphone display to grayscale[#14] led to a reduction in total phone use (d = 0.39), social media use (d = 0.44), and internet browser use (d = 0.46). No effect on video player use was found

Removing the Facebook newsfeed[#16b] resulted in a reduction for the duration of an individual session (d = 0.75). No effects were found for the total time spent on the distraction or the starts of the distracting activity.

4.2.5 Reward Intervention Effects

When social support features were added to a goal-advancement intervention[#2a], participants set significantly more time-limit goals per day (d = 1.87), reduced their time on the device (d = 0.83), and started distractions less often (d = 0.56). These changes were not observed in the variant without the social support features[#2b].

4.2.6 Effects of interventions with mixed feature contribution

For some of the interventions with multiple features, the effect of the intervention cannot be assigned to a single feature. The browser redesign#1 achieved reductions for the time on distraction and the start of distractions. However, during the intervention period there were fewer URLs total visited (d = 1.39), and fewer total tabs open in general, indicating that there was an overall reduction in activity. Rotating different behavior change interventions led to a reduction in time spent on distraction[#9a]. The rotation of features was also the focus of the most thorough investigation of intervention use. Kovacs et al. (2018) found that rotating the active interventions led to less time on distraction, but also to more attrition, that is, uninstalling the intervention[#9a]. The additional information and the additional options[#9b] that were granted to participants in the follow-up study reduced the attrition significantly. While in the control condition, only 44% of the participants remained after 7 days, the additional information led to 79% survival, and the additional options condition to 80% survival.

4.3 Study quality

In this section, we present our assessment of the study quality and the confidence that we have in the results of the studies. We start with the sampling criteria and the sample size, followed by the study duration,

For the study sample we found several recruiting criteria that were mentioned in the publications (see column 'Selection Criteria' in Table 6). These included participants with an interest in reducing smartphone usage, participants that felt that they were often distracted, or interested in improving productivity. In the study (Lottridge et al., 2012), participants were pre-screened with the Media Multitasking Index (Ophir et al., 2009, p. 20), and only those who scored either high or low in this index were included. Kim, Jung, et al. (2019) selected participants according to their readiness within the Transtheoretical Model framework (Prochaska & Velicer, 1997). None of the studies reported blinding either the participants or the investigators to the conditions. It should be acknowledged that a blinding of the participants is often not possible due to the nature of the interventions.

Table 6

Study data

ID	Duration			Design	n	Recruitment Criteria
	Pre	At	Post			
#1	7	7	-	WI	14	Undergrad students; high or low multitaskers.
#2a #2b	7	14	-	WI WI	35 27	Interest in "improving" smartphone use.
#3a #3b	7	35	-	WI	19	Willingness to decrease smartphone use.
#4	7	7	-	WI	23	Android smartphone users
#5a #5b	10	20	10	WI	12 12	Interest in self-tracking and productivity enhancement.
#6	-	7	-	BTW & WI	104	University students
#7a	2	2	-	WI	61	Office workers and students
#7b	2	2	-	BTW & WI	55	Students from a computer mediated communication class
#8	7	14	-	BTW	36	University students who "were willing to improve their productivity”
#9a	32	32	-	WI	21	Online users
#9b	10	10	-	WI	93	Online users
#10a #10b	7	7	7	WI WI	19 15	Crowdworkers Crowdworkers
#11a #11b #11c	7	7	-	WI	36	Students who want to reduce smartphone use
#12a #12b #12c	7	21	-	WI	44	Students who want to reduce excessive smartphone use
#13a #13b	5	10	-	WI	15	Office workers

#14	-	10	-	BTW &WI	161	Undergrad students
#15	21	30	-	BTW	73	Smartphone users
#16a	14	14	14	BTW	39	Students often distracted by Facebook
#16b					38	

Note. Study duration in days for the baseline measurement (Pre), the active period (At) and the post-intervention period (Post). Whether the study design was a between subjects (BTW) design or a within subjects (WI) design. The table also contains the sample size (n), and the criteria that were used for participant recruitment. Cells with one row for multiple interventions indicate that both interventions were used in the same study.

The sample sizes ranged from 12 to 217 participants (M = 52.18, SD = 51.76), and seven studies were conducted with less than 20 participants. Although there are no fixed criteria for sample size requirements, studies with participants in the range of 20 or below are considered adequate to identify usability issues for usability research purposes (Caine, 2016), but these numbers are small compared to behaviour change interventions from other domains (Norman et al., 2007).

The most frequent duration that an intervention was active for was 7 days, with a mean duration of 14.6 days (SD = 9.60). For six interventions2 we found reports about the post-intervention period.

For 10 interventions, the reporting of results did not allow a conversion into effect sizes.3 In seven of these cases, this conversion was not possible due to data that was not reported.4 For the interventions #11a, #11b, and #11c, the measure was prevented app openings, and, since the baseline period had zero prevented openings, no comparison with a baseline was possible.

One publication made their data publicly available via the servers of the Open Science Foundation (Lyngs et al., 2020).

5 Discussion

In this review, we summarized and categorized the existing evidence for digital self-control interventions. We categorized the interventions according to their features into the categories awareness, goal-advancement, blocking, feature modification, and reward and punishment. Awareness features, which were present in 16 interventions, were the most frequent. Goal setting features were present in 10 interventions, mostly represented by time-based goal setting. Nine interventions had features to block access to distracting content. Five interventions modified the content of websites, and two variants of a single intervention used a reward scheme.

To address the research question regarding the effectiveness of the interventions, we first categorized the outcomes into the categories of time spent on distractions, the start of distractions, time spent on the device, and use of the intervention itself. While positive outcomes were reported for all types of interventions, the mixture of features within some of the interventions often made it difficult to pinpoint effects of individual features. This is especially true for features from the awareness category, such as usage statistics, which are often present alongside other features. We assume that the awareness features are so clearly over-represented because it is comparatively easy to implement them. Information about time distribution on the device has to be collected for practically every intervention, one can create pretty visualizations out of these statistics, and it is thus tempting to include them. Presumably, these features do no harm, but they make it difficult to establish the degree to which each feature contributed to the effectiveness of an intervention. Nevertheless, several observations from the results stand out and should be considered for future interventions.

5.1 Awareness and Self-Monitoring are Insufficient

The results from those interventions which only contained awareness features indicate that relying solely on the users monitoring and adapting their own behaviour is not effective. This reflects the insight that people often avoid monitoring themselves, especially regarding uncomfortable topics (Chang et al., 2017). Moreover, several studies also reported greater effects for the beginning of their intervention, which suggests that a novelty effect could be at least partially responsible for the effectiveness of interventions (Tsay et al., 2019), while habitual behaviour may gain the upper hand again in the future. Whether it is through deliberate ignoring or habituation, it appears that interventions that users can easily ignore do not achieve the desired results.

5.2 Circumvention Difficulty Can Tip the Scales

On the other hand, interventions that are more insistent in grabbing the users' attention at times of excessive distraction use are more effective. The insistence with which an intervention attempts to convince the users, and the ease with which participants can dismiss interventions appear to be relevant factors for the success of interventions. Against the background of autonomy needs and reactance (Ryan & Deci, 2000), it makes sense to give users a way to influence the strictness of interventions, and the results of those studies where users have a choice confirm a preference for interventions that allow negotiation. At the same time, too little resistance to bypassing sanctions once again leads to ignoring the interventions completely. This balancing act requires the ability to have adaptive sanctioning strictness (Schwartz et al., 2021).

5.3 Detection of Context

Although the majority of the interventions were tested on students, none of the included studies was conducted explicitly in a learning context. To confirm the applicability in learning situations, future interventions should take the context into consideration. A simple way to do this would be to ask the users what they are currently doing. A less interruptive approach could be to monitor learning activity in digital learning environments, and infer active learning from there.

However, even when the learning context is known, distracting content cannot always be identified with certainty, when we consider learners that may need to communicate with their peers on a social media site or watch learning-related content on the same platform where they watch entertaining videos. Thus, in addition to reliable context detection, the classification into distracting or learning relevant content needs to become more sophisticated, beyond simple URL- or app-based blacklists.

5.4 Interventions for Specific Groups

All the interventions were used with student or worker populations, and none with the especially susceptible group of adolescents (van der Schuur et al., 2015). Considering that this is an especially at-risk population, they should be the target of future research (E. J. Lee & Ogbolu, 2018).

Similarly, there are differences in traits between people that may predispose them to respond better or worse to specific features (Mark et al., 2018). Those factors have to be identified for adaptive interventions.

5.5 Long-Term Effects

Another topic that needs to be addressed is how well the results from an intervention transfer to longer lasting behaviour changes. Those interventions in this review that address this question with follow-ups showed no longer lasting changes.

5.6 Limitations

One limitation that is inherent to reviews is that publications are missed due to the inclusion criteria. Many of these interventions, such as group-based locking of smartphones (Kim, Jung, et al., 2017), using smartwatches (Dibia, 2016) or a physical doll that reminds of excessive usage (Choi et al., 2016), seemed promising or used innovative approaches. We also did not incorporate interviews and other forms of qualitative analysis in this review, which often provide more nuances regarding the intervention and its reception. For those interested in a particular kind of intervention, these additional insights should be considered.

6 Conclusion

There is no reason to assume that digital content will become less distractive, and that the attempts to grab our attention at every waking moment will lessen. Digital self-control interventions are one of the ways to address this problem, but there remain a lot of open issues before they can be called a solution.

References

Baumgartner, S. E., Lemmens, J. S., Weeda, W. D., & Huizinga, M. (2017). Measuring Media Multitasking: Development of a Short Measure of Media Multitasking for Adolescents. *Journal of Media Psychology*, *29*(2), 1–10. https://doi.org/10.1027/1864-1105/a000167

Bayer, J. B., & LaRose, R. (2018). Technology Habits: Progress, Problems, and Prospects. In B. Verplanken (Ed.), *The Psychology of Habit* (pp. 111–130). Springer International Publishing. https://doi.org/10.1007/978-3-319-97529-0_7

Caine, K. (2016). Local Standards for Sample Size at CHI. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 981–992. https://doi.org/10.1145/2858036.2858498

Chang, B. P. I., Webb, T. L., & Benn, Y. (2017). Why Do People Act Like the Proverbial Ostrich? Investigating the Reasons That People Provide for Not Monitoring Their Goal Progress. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00152

Choi, S., Jeong, H., Ko, M., & Lee, U. (2016). LockDoll: Providing Ambient Feedback of Smartphone Usage within Social Interaction. *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '16*, 1165–1172. https://doi.org/10.1145/2851581.2892445

Coyne, M. A., & Wright, J. P. (2014). The stability of self-control across childhood. *Personality and Individual Differences*, *69*, 144–149. https://doi.org/10.1016/j.paid.2014.05.026

Dibia, V. (2016). FOQUS: A Smartwatch Application for Individuals with ADHD and Mental Health Challenges. *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*, 311–312. https://doi.org/10.1145/2982142.2982207

Duckworth, A. L., Taxer, J. L., Eskreis-Winkler, L., Galla, B. M., & Gross, J. J. (2019). Self-Control and Academic Achievement. *Annual Review of Psychology*, *70*(1), 373–399. https://doi.org/10.1146/annurev-psych-010418-103230

Eyal, N. (2014). *Hooked: How to build habit-forming products*. Portfolio/Penguin.

Fiorella, L. (2020). The Science of Habit and Its Implications for Student Learning and Well-being. *Educational Psychology Review*. https://doi.org/10.1007/s10648-020-09525-1

Foulonneau, A., Calvary, G., & Villain, E. (2016). Stop procrastinating: TILT, time is life time, a persuasive application. *Proceedings of the 28th Australian Conference on Computer-Human Interaction - OzCHI '16*, 508–516. https://doi.org/10.1145/3010915.3010947

Hiniker, A., Hong, S. (Ray), Kohno, T., & Kientz, J. A. (2016). MyTime: Designing and Evaluating an Intervention for Smartphone Non-Use. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 4746–4757. https://doi.org/10.1145/2858036.2858403

Holte, A. J., & Ferraro, F. R. (2020). True colors: Grayscale setting reduces screen time in college students. *The Social Science Journal*, 1–17. https://doi.org/10.1080/03623319.2020.1737461

Kim, J., Cho, C., & Lee, U. (2017). Technology Supported Behavior Restriction for Mitigating Self-Interruptions in Multi-device Environments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *1*(3), 64:1-64:21. https://doi.org/10.1145/3130932

Kim, J., Jung, H., Ko, M., & Lee, U. (2019). GoalKeeper: Exploring Interaction Lockout Mechanisms for Regulating Smartphone Use. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *3*(1), 16:1-16:29. https://doi.org/10.1145/3314403

Kim, J., Park, J., Lee, H., Ko, M., & Lee, U. (2019). LocknType: Lockout Task Intervention for Discouraging Smartphone App Use. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, 1–12. https://doi.org/10.1145/3290605.3300927

Kim, Y.-H., Jeon, J. H., Choe, E. K., Lee, B., Kim, K., & Seo, J. (2016). TimeAware: Leveraging Framing Effects to Enhance Personal Productivity. *Proceedings of the*

2016 CHI Conference on Human Factors in Computing Systems, 272–283. https://doi.org/10.1145/2858036.2858428

Kirschner, P. A., & Karpinski, A. C. (2010). Facebook® and academic performance. *Computers in Human Behavior*, *26*(6), 1237–1245. https://doi.org/10.1016/j.chb.2010.03.024

Ko, M., Yang, S., Lee, J., Heizmann, C., Jeong, J., Lee, U., Shin, D., Yatani, K., Song, J., & Chung, K.-M. (2015). NUGU: A Group-based Intervention App for Improving Self-Regulation of Limiting Smartphone Use. *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 1235–1245. https://doi.org/10.1145/2675133.2675244

Kovacs, G., Wu, Z., & Bernstein, M. S. (2018). Rotating Online Behavior Change Interventions Increases Effectiveness But Also Increases Attrition. *Proceedings of the ACM on Human-Computer Interaction*, *2*(CSCW), 1–25. https://doi.org/10.1145/3274364

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00863

Lee, E. J., & Ogbolu, Y. (2018). Does Parental Control Work With Smartphone Addiction?: A Cross-Sectional Study of Children in South Korea. *Journal of Addictions Nursing*, *29*(2), 128–138. https://doi.org/10.1097/JAN.0000000000000222

Lee, H., Ahn, H., Choi, S., & Choi, W. (2014). The SAMS: Smartphone Addiction Management System and Verification. *Journal of Medical Systems*, *38*(1). https://doi.org/10.1007/s10916-013-0001-1

Loid, K., Täht, K., & Rozgonjuk, D. (2020). Do pop-up notifications regarding smartphone use decrease screen time, phone checking behavior, and self-reported problematic smartphone use? Evidence from a two-month experimental study. *Computers in Human Behavior*, *102*, 22–30. https://doi.org/10.1016/j.chb.2019.08.007

Lottridge, D., Marschner, E., Wang, E., Romanovsky, M., & Nass, C. (2012). Browser Design Impacts Multitasking. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *56*(1), 1957–1961. https://doi.org/10.1177/1071181312561289

Lyngs, U., Lukoff, K., Slovak, P., Binns, R., Slack, A., Inzlicht, M., Van Kleek, M., & Shadbolt, N. (2019). Self-Control in Cyberspace: Applying Dual Systems Theory to a Review of Digital Self-Control Tools. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–18. https://doi.org/10.1145/3290605.3300361

Lyngs, U., Lukoff, K., Slovak, P., Seymour, W., Webb, H., Jirotka, M., Van Kleek, M., & Shadbolt, N. (2020). "I Just Want to Hack Myself to Not Get Distracted": Evaluating Design Interventions for Self-Control on Facebook. *ArXiv:2001.04180 [Cs]*. https://doi.org/10.1145/3313831.3376672

Mark, G., Czerwinski, M., & Iqbal, S. T. (2018). Effects of Individual Differences in Blocking Workplace Distractions. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, 1–12. https://doi.org/10.1145/3173574.3173666

Masood, A., Luqman, A., Feng, Y., & Ali, A. (2020). Adverse consequences of excessive social networking site use on academic performance: Explaining underlying mechanism from stress perspective. *Computers in Human Behavior*, *113*, 106476. https://doi.org/10.1016/j.chb.2020.106476

Mazar, A., & Wood, W. (2018). Defining Habit in Psychology. In B. Verplanken (Ed.), *The Psychology of Habit* (pp. 13–29). Springer International Publishing. https://doi.org/10.1007/978-3-319-97529-0_2

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Medicine*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Norman, G. J., Zabinski, M. F., Adams, M. A., Rosenberg, D. E., Yaroch, A. L., & Atienza, A. A. (2007). A Review of eHealth Interventions for Physical Activity and Dietary Behavior Change. *American Journal of Preventive Medicine*, *33*(4), 336-345.e16. https://doi.org/10.1016/j.amepre.2007.05.007

Okeke, F., Sobolev, M., Dell, N., & Estrin, D. (2018). Good vibrations: Can a digital nudge reduce digital overload? *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '18*, 1–12. https://doi.org/10.1145/3229434.3229463

Ophir, E., Nass, C., & Wagner, A. D. (2009). Cognitive control in media multitaskers. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(37), 15583–15587. https://doi.org/10.1073/pnas.0903620106

Parry, D. A., & le Roux, D. B. (2019). Media multitasking and cognitive control: A systematic review of interventions. *Computers in Human Behavior*, *92*, 316–327. https://doi.org/10.1016/j.chb.2018.11.031

Petticrew, M., & Roberts, H. (Eds.). (2006). *Systematic Reviews in the Social Sciences*. Blackwell Publishing Ltd. https://doi.org/10.1002/9780470754887

Prochaska, J. O., & Velicer, W. F. (1997). The transtheoretical model of health behavior change. *American Journal of Health Promotion: AJHP*, *12*(1), 38–48. https://doi.org/10.4278/0890-1171-12.1.38

Rosen, L. D., Mark Carrier, L., & Cheever, N. A. (2013). Facebook and texting made me do it: Media-induced task-switching while studying. *Computers in Human Behavior*, *29*(3), 948–958. https://doi.org/10.1016/j.chb.2012.12.001

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Schwartz, R. X., Monge Roffarello, A., De Russis, L., & Apostolellis, P. (2021). Reducing Risk in Digital Self-Control Tools: Design Patterns and Prototype. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–7. https://doi.org/10.1145/3411763.3451843

Shea, N., Boldt, A., Bang, D., Yeung, N., Heyes, C., & Frith, C. D. (2014). Supra-personal cognitive control and metacognition. *Trends in Cognitive Sciences*, *18*(4), 186–193. https://doi.org/10.1016/j.tics.2014.01.006

Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron*, *79*(2), 217–240. https://doi.org/10.1016/j.neuron.2013.07.007

Terry, C. A., Mishra, P., & Roseth, C. J. (2016). Preference for multitasking, technological dependency, student metacognition, & pervasive technology use: An experimental intervention. *Computers in Human Behavior*, *65*, 241–251. https://doi.org/10.1016/j.chb.2016.08.009

Tsay, C. H.-H., Kofinas, A. K., Trivedi, S. K., & Yang, Y. (2019). Overcoming the novelty effect in online gamified learning systems: An empirical evaluation of student engagement and performance. *Journal of Computer Assisted Learning*, *n/a*(n/a). https://doi.org/10.1111/jcal.12385

Tseng, V. W.-S., Lee, M. L., Denoue, L., & Avrahami, D. (2019). Overcoming Distractions during Transitions from Break to Work using a Conversational Website-Blocking System. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, 1–13. https://doi.org/10.1145/3290605.3300697

van der Schuur, W. A., Baumgartner, S. E., Sumter, S. R., & Valkenburg, P. M. (2015). The consequences of media multitasking for youth: A review. *Computers in Human Behavior*, *53*, 204–215. https://doi.org/10.1016/j.chb.2015.06.035

Varol, F., & Yıldırım, E. (2019). Cyberloafing in Higher Education: Reasons and Suggestions from Students' Perspectives. *Technology, Knowledge and Learning*, *24*(1), 129–142. https://doi.org/10.1007/s10758-017-9340-1

Whittaker, S., Kalnikaite, V., Hollis, V., & Guydish, A. (2016). "Don't Waste My Time": Use of Time Information Improves Focus. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 1729–1738. https://doi.org/10.1145/2858036.2858193

Willems, Y. E., Boesen, N., Li, J., Finkenauer, C., & Bartels, M. (2019). The heritability of self-control: A meta-analysis. *Neuroscience and Biobehavioral Reviews*, *100*, 324–334. https://doi.org/10.1016/j.neubiorev.2019.02.012

Wood, W., & Rünger, D. (2016). Psychology of Habit. *Annual Review of Psychology*, *67*(1), 289–314. https://doi.org/10.1146/annurev-psych-122414-033417

Paper 4

Biedermann, D., Kister, S., Breitwieser, J., Weidlich, J., and Drachsler, H. (2023). Use of Digital Self-Control Tools in Higher Education – A Survey Study. [This paper is in its second revision in the journal *Education and Information Technology*]

Use of Digital Self-Control Tools in Higher Education – A Survey Study

Daniel Biedermann[1], Stella Kister[2], Jasmin Breitwieser[2], Joshua Weidlich[1], and Hendrik Drachsler[1,3]

[1] Information Center for Education, DIPF | Leibniz Institute for Research and Information in Education, Frankfurt am Main, Germany

[2] Education and Human Development, DIPF | Leibniz Institute for Research and Information in Education, Frankfurt am Main, Germany

[3] Computer Science Faculty, Goethe University Frankfurt, Frankfurt am Main, Germany

[4] Institute of Psychology, Goethe University Frankfurt, Frankfurt am Main, Germany

Author Note

Daniel Biedermann https://orcid.org/0000-0001-9219-222X

Stella Kister https://orcid.org/0000-0001-8871-9375

Jasmin Breitwieser https://orcid.org/0000-0002-0337-7159

Joshua Weidlich https://orcid.org/0000-0002-1926-5127

Hendrik Drachsler https://orcid.org/0000-0001-8407-5314

Corresponding Author

Correspondence concerning this article should be addressed to Daniel Biedermann, Information Center for Education, DIPF | Leibniz Institute for Research and Information in Education, Rostocker Str. 6, 60323 Frankfurt am Main, Germany. E-mail: d.biedermann@dipf.de, phone: +49 69 24708 173

Other Author Contact

Jasmin Breitwieser, Education and Human Development, DIPF | Leibniz Institute for Research and Information in Education, Rostocker Str. 6, 60323 Frankfurt am Main, Germany.
E-mail: j.breitwieser@dipf.de, phone: +49 69 24708 375

Stella Kister, Goethe-University Frankfurt am Main
E-mail: stella.kister@gmail.com

Joshua Weidlich, Information Center for Education, DIPF | Leibniz Institute for Research and Information in Education, Rostocker Str. 6, 60323 Frankfurt am Main, Germany.
E-mail: j.weidlich@dipf.de, phone: +49 69 24708 870

Hendrik Drachsler, Information Center for Education, DIPF | Leibniz Institute for Research and Information in Education, Rostocker Str. 6, 60323 Frankfurt am Main, Germany.
E-mail: h.drachsler@dipf.de, phone: +49 69 24708 870

Abstract

Distractions are ubiquitous in today's technology-saturated environments, an issue that significantly impacts learning contexts employing digital technologies and yields detrimental effects on learning. Digital self-control tools, which aim to assist users in their efforts to reduce digital distractions, are numerous and readily available. Despite several dedicated empirical studies focusing on specific tools, there remains a notable lack of information regarding their daily use and helpfulness. Furthermore, the sheer variety of these tools prompts questions about their universal helpfulness and the potential influence of individual differences.

To address these issues, we surveyed a sample of higher-education students, totaling 273 individuals. These students reported on their media use, satisfaction with learning, and experiences with features of digital self-control tools. Our study's findings indicate a discrepancy in the perception and awareness of these features; those deemed most helpful are among the least known, and conversely, common features are often perceived as unhelpful.

Our research also uncovered a negative correlation between habitual media use and the use of less restrictive features. Another identified issue was constraints on the use of these tools for learning, as platforms often serve dual purposes for both education and entertainment. We delve into these practical problems and propose future research directions to further advance the understanding of digital self-control tools.

Keywords

Digital self-control tools; Digital Distractions; Digital Wellbeing; self-control; Self-Regulation; Media Multitasking

1 Introduction

With the growing availability of digital technology, students are exposed to a constant stream of digital distractions, which can lead to conflicts with their learning goals. On one hand, students may be drawn to engaging in enjoyable activities, such as browsing social media or watching entertaining videos. On the other hand, students may also recognize the importance of achieving long-term academic goals, which require focus and effort. This self-control conflict between immediate gratification and delayed rewards is a growing problem that negatively impacts students' academic performance and psychological well-being (Jamet et al., 2020; Masood et al., 2020; Ratan et al., 2021; Tran et al., 2019). Studies revealed that students use digital distractions frequently, often averaging only a few minutes of uninterrupted study time before turning to digital media (Calderwood et al., 2014; Rosen et al., 2013). Nonetheless, there are tools that students can use to mitigate the negative effects of these distractions.

In recent years, digital self-control tools (DSCTs) have emerged to support the self-control of students (Biedermann et al., 2021; Lyngs et al., 2019). DSCTs come in a variety of forms. There are, for instance, website blockers that allow users to block access to digital content on their devices (e.g., J. Kim et al., 2017), visualizations that show users how much time they spend on digital content, in the hope that this will lead to a change in behaviour (e.g., (Y.-H. Kim et al., 2016), or reminders that pop up when a user spends too much time on digital distractions (e.g., J. Kim, Jung, et al., 2019). Aside from these very common examples, there are many more such tools with a variety of ingenious features to support self-control and mitigate distractions (Lyngs et al., 2019). Some of these we will present in the latter sections of this paper, and we encourage interested readers to visit the app stores of their preferred device to view the diversity of tools for themselves.

Indeed, most digital devices now come with pre-installed DSCTs, such as the "Digital Wellbeing" app on Android devices, or the "Focus" app on iOS devices. The investment of device manufacturers in tools to reduce device use, despite their vested interest in the opposite, underscores that digital distractions have been recognized as a severe problem. A yet unresolved question is whether DSCTs actually address this problem effectively, especially to alleviate their negative effects on academic achievement (Biedermann et al., 2021).

The solution appears obvious, a student who suffers from frequently watching videos instead of learning could install a website blocker. The content is no longer accessible, the problem should be solved. Why does this not happen whenever a student suffers from digital distractions?

As the first possible obstacle to the widespread use of self-control tools, we suspect that they may simply not be well known enough. There are all kinds of helpful things that people do not know about, and we assume that DSCTs are no exception. Even the pre-installed apps are not activated by default, and users must discover them on their devices. To our knowledge, there has been no research to date on how widespread the use of DSCTs is.

A second barrier may be found in a lack of motivation. Students might be aware of DSCTs, but not be sufficiently motivated to use them to limit digital distractions. They might recognize that distractions are an issue but feel that watching videos is simply too enjoyable, and a website blocker would prevent this enjoyment. Perhaps the deadline for the essay is several weeks in the future.

The next hurdle is selecting the right DSCT. Due to differences in the way a particular feature is implemented, and due to interindividual differences in how one responds to a particular feature, not every person benefits to the same extent from the same DSCT.

2 The Role of Differences in DSCTs

2.1 Differences Due to Implementation Details

Tools that nominally have the same mechanism can implement their self-control features very differently. This, of course, affects how users perceive the tool, because expectations of effectiveness and ease of use play a big role in the adoption of technologies (Granić & Marangunić, 2019; Nguyen, 2022). For example, the success of content blocking is related to the ease with which the blocking can be disabled. When users had to enter a code of random digits before they could lift a block, the effectiveness was significantly higher for longer codes. At the same time, the longer codes also involved more effort, and therefore tended to be less popular (J. Kim, Park, et al., 2019). Similar observations were made when users set time limits for device usage. It worked well only in conjunction with automatic lockouts from the device after the time limit. If the users could extend their limits freely, they mostly did just that and clicked their reminders away (J. Kim, Jung, et al., 2019). In general, DSCTs which do not apply any restrictions, and rely purely on users monitoring themselves, tend to be ineffective (Loid et al., 2020; Terry et al., 2016; Zimmermann, 2021).

2.2 Differences in Habitual Media Use

Of course, there is a plethora of individual differences that may lead to differential preferences for a DSCT. But the actual effectiveness, independent of preference, could be influenced by the degree of which the media use is habitual. Habits are automatic and unconscious behaviours triggered by environmental cues that bypass conscious, goal-oriented decision-making processes (Wood & Rünger, 2016). This automaticity can render certain DSCTs, which require cognitive effort and conscious decision-making, less effective. Following the taxonomy of Lyngs et al. (2019), DSCTs with self-tracking and goal advancement features work by comparing current behaviour with specified goals, and thus implies that conscious decisions about goals are made. On the other hand, preventing features like blocking should be better suited to prevent unwanted habitual behaviour because they are active even when there is no conscious goal driving the behaviour.

2.3 Challenges Due to Dual Use of Media

We also see a special challenge when using DSCTs while learning: The same platforms can be both relevant for on- and off-task purposes. YouTube has a lot of entertaining videos, but it also has lecture videos and a lot of high-quality educational content. Similarly, a social network might be the source of interruptions, but also the place where other students exchange essential information about a course (Hrastinski & Aghaee, 2012). Thus, completely preventing access is not always a viable option. Users have to micro-manage and activate or deactivate their DSCT whenever they start or stop learning, possibly even depending on the specific task they work on. Micro-managing requires the user to switch from their original task to the task of managing their tool. Task switching requires effort and can result in poorer performance of the original task (Rubinstein et al., 2001). Consequently, it appears that users will start to circumvent or even disable restrictive mechanisms frequently (c.f. (J. Kim, Park, et al., 2019)).

In the previous sections, we have sketched out a scenario that is as follows: On the one hand, there are multiple interventions, some of which have also been shown to be effective in a number of studies (Holte & Ferraro, 2020; J. Kim, Park, et al., 2019; Ko et al., 2015; Lyngs et al., 2020; Tseng et al., 2019). On the other hand, we see that digital distractions continue to be a problem that DSCTs seem to not fully address. We have gone through several possible reasons for this: Insufficient awareness, deficiencies in the implementation, and a poor fit between the individual and the chosen tool.

In this study, we examine whether and to what extent these assumptions hold true in a sample of higher education students.

3 The Present Study

We administered a survey to investigate the prevalence of different DSCT features among students, their knowledge about them, and the perceived helpfulness of DSCT features for reducing digital distractions during learning. We focus on higher education students because they are typically of full age and, therefore, rather free in their time management, with less external regulation. Even in university classrooms, media use is typically unregulated (Wammes et al., 2019). It can be assumed that DSCT use among higher education students is voluntary and self-determined. Of course, DSCTs can also have a benefit for school children, but the environment is fundamentally different, with more external regulation from school policies (Tandon et al., 2020) and from parents (Nikken & De Haan, 2015).

We tailored our questions to the perception of individual features (e.g., website blocking, goal setting) rather than specific tools. This was to avoid the potential confounding influence of multifaceted tools, which typically combine numerous features such as website blocking, usage visualization, and goal-prompting within a single application (Lyngs et al., 2019). This has the benefit of generating potential insights that are not tied to specific applications and instead shed light on the underlying mechanisms and affordances associated with different features (Ko et al., 2015; Kovacs et al., 2018).

First, we wanted to rule out that DSCTs fail simply because users are not aware that they exist. Put differently, if users do not know about a solution, then it cannot help them. Thus, our first research question (RQ) addressed the knowledge about and actual use of DSCTs.

RQ1-Knowledge & Use: How widespread are knowledge about and use of DSCT features?

Building on this, our second research question aimed to investigate whether participants perceived DSCT features as actually helpful in mitigating the negative impact of digital distractions.

RQ2-Helpfulness: How do participants perceive the helpfulness of DSCT features in reducing digital distractions during learning?

Further, we examine the connection between habitual use of media and the helpfulness of DSCTs features. Habitual behaviour bypasses goal-directed behaviour, and we therefore suspect that DSCTs which rely on an individual's assessment of conscious goals are less useful in these cases (Chang et al., 2017; Pinder et al., 2018).

RQ3-Habits: How are the perceived helpfulness of DSCT features and habitual use of digital media correlated?

There are indications that users are often frustrated with the DSCTs that they use, and that these frustrations may cause users to stop using an otherwise helpful tool. (Lyngs et al., 2022). We would therefore like to learn more about why users stop using the tool after they have already used it.

RQ4-Stopping: What are the reasons that people have for stopping the use of a DSCT?

4 Methods

To investigate our research questions, we employed a correlational survey design, sampling a wide variety of people to account for the breadth of experiences with and attitudes toward DSCTs. In addition to a quantitative section of our survey, we also gathered more qualitative information via free text responses. As recruitment procedure, we reached participants via social media, word of mouth, mailing lists, and the prolific recruitment platform.

4.1 Participants

403 participants completed the questionnaire over a period of three months between May and August 2022. Ethical approval for the study was obtained from the ethics committee of [blinded for peer review]. All participants had to consent to the use of their data for the study prior to starting the questionnaire.

We excluded all participants that did not complete the full questionnaire (n = 46). Due to our focus on the use of DSCTs for learning scenarios, we excluded all participants that were not enrolled students (university, school, or vocational) or Ph.D. candidates (n = 132). To ensure data quality, the questionnaire contained a screener question. Participants that did not respond to this question as instructed were excluded from the analysis (n = 35). The final number of participants was 273. The mean age of the participants was 27.18. (SD = 7.56), 84 male, 186 female, three diverse. Countries of residence were 141 Germany, 77 UK, 24 US, 10 Ireland, 4 Australia, 4 Austria, 3 Italy, 3 Netherlands, 2 Portugal, 1 France, 1 Philippines, 1 Norway, 1 Greece, and 1 Columbia. Regarding their occupation, 219 responded that they were enrolled students (202 university, 17 vocational), and 54 were Ph.D. candidates.

4.2 Instruments

4.2.1 Questions About the Extent and Effect of Digital Distractions During Learning

We used one item to ask about the frequency of digital distractions in general (scale from 1 (never) to 5 (very often)), one item to ask whether digital distractions are a hindrance to reaching their learning goals (scale from 1 (does not apply) to 4 (applies)), one item to ask about suffering from digital distractions (scale from 1 (does not apply) to 4 (applies)), and one item for the self-efficacy of being able to reduce digital distractions during learning (scale from 1 (does not apply) to 4 (applies)). Satisfaction with learning outcomes was assessed with two items (scale from 1 (dissatisfied) to 4 (satisfied)) (e.g., "In general, how satisfied are you with the results that you get from your learning?", Cronbach's $\alpha = 0.74$).

4.2.2 Habituality of Media Use

We measured the degree of habitual media use with a modified version of the self-report habit index (SRHI) developed by Verplanken and Orbell (Verplanken & Orbell, 2003). The SRHI is a widely used measure of habit strength and has been shown to have good reliability and validity in previous research (Gardner et al., 2012; Verplanken & Orbell, 2003). We used only the four items that, according to Gardner et al. (2012), can measure the automaticity of habits (termed the "Self-Report Behavioural Automaticity Index"; SRBAI). These items are: '[Behaviour X is something…]' '...I do automatically', '...I do without having to consciously remember', '...I do without thinking', and '...I start doing before I realise I'm doing it'. Each of these items was rated on a Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree).

We used these items to measure the habitual use of the following behaviours: using social media ("Using social media while or instead of learning is something that ...", $\alpha =$ 0.91), watching videos ("Watching videos or series while or instead of learning is something that ...", $\alpha = 0.92$), surfing the internet ("Surfing the internet while or instead of learning is something that ...", $\alpha = 0.92$), gaming ("Playing video games (e.g. on PC, console, or smartphone) while or instead of learning is something that ...", $\alpha = 0.89$), and chatting ("Communicating with others via messenger services while or instead of learning is something that...", $\alpha = 0.89$). To capture the presence of habitual media-related distractions in general, we report the highest SRBAI score across all SRBAI scores per participant (MaxSRBAI).

4.2.3 Knowledge and Perceived Helpfulness of DSCTs

Knowledge of and prior experience with DSCT features were rated on a per-feature basis, with questions about one feature on one page. We composed the list of features based on existing DSCT research (Biedermann et al., 2021; Lyngs et al., 2019; Roffarello & De Russis, 2022), personal experience, and feedback from pilot studies.

- Autoclose: A feature that automatically closes apps or websites after a specified amount of time.
- Blocking: A feature that blocks access to specific apps or websites
- Delay: A feature that uses delay of gratification mechanisms, such as making the user wait or solve a task before accessing an app or website.
- Modification: A feature that removes or modifies particularly distracting features from websites.
- Gamification: A feature that uses game-like elements, such as rewards, to motivate users to engage in less distracting behaviour.
- Goals: A feature that prompts users to create and track progress towards specific goals,
- Pomodoro: A feature that supports the use of the Pomodoro technique, where the user specifies time intervals (typically 25 minutes) that are reserved for focused work, after which they are allowed a short break (typically 5 minutes) to do whatever they like.
- Compare: A feature that enables users to share and compare their progress with others, which can help users stay motivated and accountable.
- Screenshare: A feature that allows users to create learning groups and monitor each other's device use to encourage less distraction during learning.
- Visualizations: A feature that provides visualizations of device usage to help users monitor and reflect on their own behaviour.

For each of these features, we created a written description and an example of a tool that has this feature at its core. We asked the participants whether they were aware that the feature exists and whether they currently use it or have used it in the past. We asked whether they think that this feature is helpful to reduce digital distractions on a scale from 1 (does not apply) to 5 (applies). We used perceived helpfulness here because it became apparent in the piloting of our questionnaire that participants found terms like effectiveness more difficult to assess than the subjective perceived helpfulness.

If a participant reported that they had previously used a DSCT but stopped doing so, or if they stated that they used a DSCT feature less than they used to, we asked them to elaborate on the reasons in a free text field. The name of the tool with the feature the participants used had to be entered in a text field.

4.2.4 Coding of Free Text Responses for Stopping Tool Use

The free-text responses about reasons for stopping tool use were first coded by one author in a round of open coding by reading all responses and creating an initial coding scheme. The remaining authors then evaluated the initial scheme by independently applying the codes to 50 responses. After discussing modifications to the coding scheme, the first author coded all the responses with the revised coding scheme, and the other authors each coded half of the responses. Pairwise intercoder agreements were 85.82% and 77.42% (M = 81.62%). The remaining discrepancies were resolved through group discussion until a consensus was reached.

In total, the participants gave 382 responses. Six responses were found to be incomprehensible, and 21 either misread the question or responded to a different question, bringing the number of coded responses to 355. A response could have multiple codes. We used the MaxQDA software for the coding process.

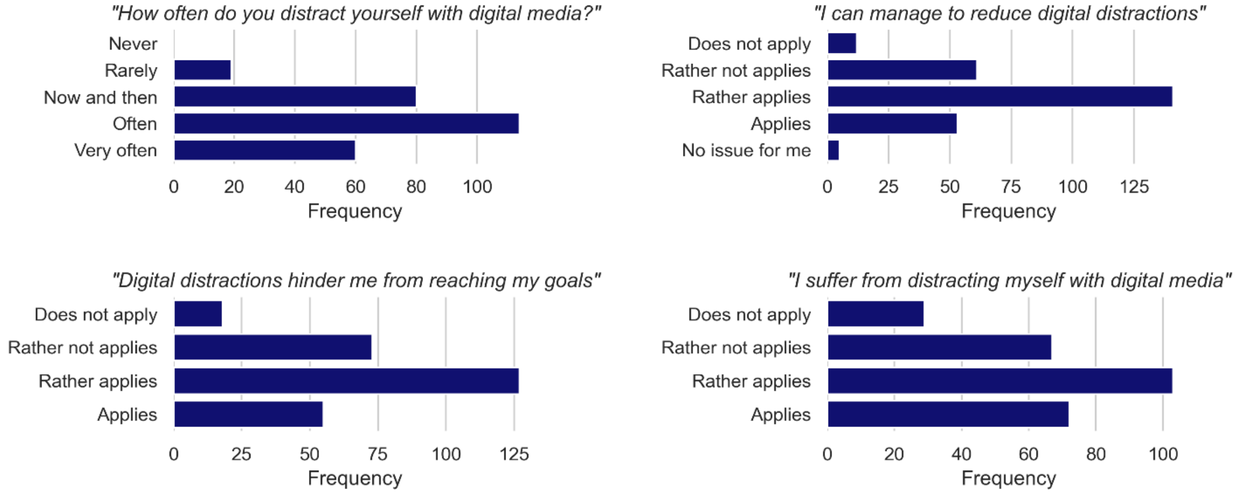
5 Results

5.1 Extent and Effect of Digital Distractions During Learning

Regarding the frequency of digital distractions, no participant stated that they never distracted themselves with digital media. Meanwhile, 19 (6.96%) said that they do it rarely, 80 (29.30%) now and then, 114 (41.75%) often, and 60 (21.98%) very often. For the question of whether the participants suffered from the digital distractions, the average rating was 2.80 (SD = 0.95) out of 5. Goal hindrance was rated similarly, with a mean of 2.80 (SD = 0.83). However, most people thought they could manage to reduce their digital distractions, and the mean rating for the self-efficacy question was 2.83 (SD = 0.86) (see Figure 1).

Figure 4

Distribution of responses to the questions regarding the extent of digital distractions during learning.



5.2 Knowledge and Use of Self-Control Tools

We explored which DSCT features the participants knew about, and which DSCT features the participants had previous experience with. For details about knowledge and usage, see Table 1. Knowledge of features varied between 6.59% (n = 18) of participants for Delay and 66.67% (n = 182) of participants for Screenshare. Prior experience (either current use or previous use that then stopped) was highest for Screenshare (n = 82, 30.04%) and particularly low for Delay (n = 5, 1.83%) and Compare (n = 5, 1.83%). Not knowing a single DSCT feature was the response of 19 (6.96%) participants, and 141 (51.65%) stated that they currently did not use any of the DSCT features that we asked about. Since it could be that this was mostly the group that did not suffer at all from distractions, we checked the sub-group of sufferers (i.e., they selected “rather applies” or “applies” on the suffering question). In this sub-group of 175 sufferers, 12 (4.40%) participants did not know any, and 86 (31.50%) participants did not currently use any DSCT features.

The specific DSCTs used by the participants can be found in the OSF repository (https://tinyurl.com/uba426xf). For brevity, we will not list all statistics here, but we noted that the goal category contained several habit trackers, to-do list apps, and calendars, which did appear to be explicitly related to reducing media use. Upon investigation, we found that the description of the goal support feature in our questionnaire allowed for the interpretation that this referred to general goal-setting tools, and not only goals related to media use.

Table 1

Statistics about tool use and knowledge.

Tool	Know	Use	Quit	Helpfulness for non-users		Helpfulness for users		Difference users and non-users	
				M	SD	M	SD	Cohen's *d*	p
Autoclose	111	31	21	3.59	0.77	3.81	1.03	0.24	.213
Blocking	108	11	31	3.68	0.96	3.88	0.97	0.21	.298
Delay	18	3	2	3.15	0.90	3.40	0.89	0.27	.609
Modification	29	5	4	3.40	0.82	4.56	0.53	1.55	.001*
Gamification	73	11	23	2.95	1.19	3.79	1.20	0.71	.004*
Goals	80	13	20	3.74	0.74	3.85	0.83	0.13	.558
Pomodoro	162	38	52	3.96	0.94	4.08	0.88	0.13	.406
Screenshare	182	39	43	3.45	1.18	4.09	0.97	0.58	.000**
Compare	21	5	0	2.19	0.83	3.20	1.48	1.01	.064
Visualization	150	56	21	3.15	1.13	3.48	1.10	0.30	.071

Note. Know: Number of participants that know about the features, Use: Number of participants that currently use a DSCT with the feature, Quit: Number and percentage of participants that have previously used the feature. * $p < .01$; ** $p < .001$.

5.3 Helpfulness of Self-Control Tools

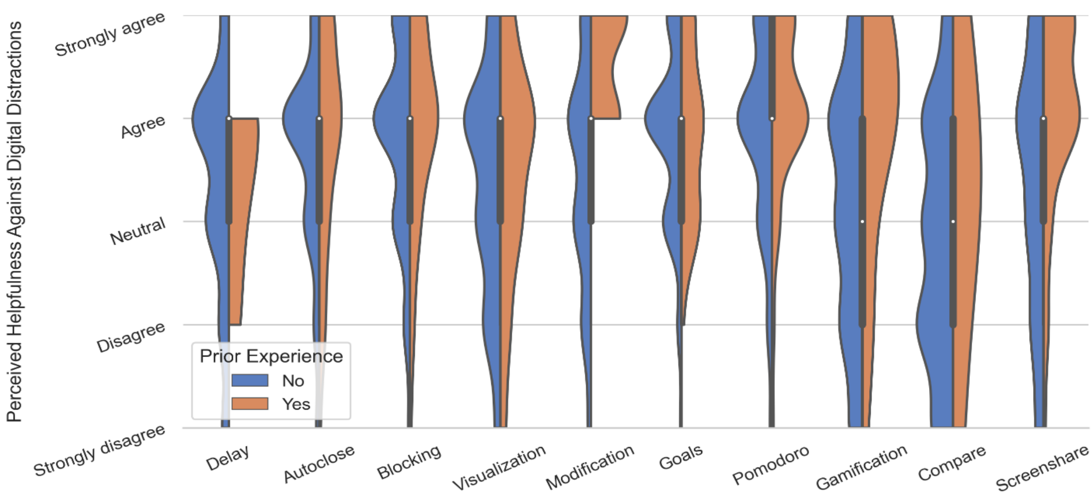
Next, we analysed the ratings for the perceived helpfulness of the DSCT features. We are primarily interested in the ratings from the participants with prior experience. Across all DSCT features, there was a trend that participants who had prior experience with a feature rated its helpfulness higher than participants without prior experience.

The mean difference between the rating from participants with prior experience and participants without prior experience (Δ) was most prominent for Gamification (Δ = 0.85, d = .71, p = 0.0035), Modification (Δ = 1.16, d = 1.55, p = 0.0006) and Screenshare (Δ = 0.64, d = 0.58, p = 0.0001). Leaving out the three features with barely any ratings, the highest rating among experienced users was for Modification (M = 4.56, SD = 0.53), and the lowest for Visualization (M = 3.48, SD = 1.10), with a mean difference between both of 1.07 (d = 1.25, p = 0.0049). See Table 1 for details regarding the perceived helpfulness of all features.

The ratings for the helpfulness of the features were skewed towards rating the features as helpful, and participants rarely rated the features as completely unhelpful (see figure 2). Even participants who stated that they stopped using a tool and explicitly stated that they did not perceive it as helpful at all rated those features as at least somewhat helpful or even very helpful.

Figure 5

Distribution of the rating for the perceived helpfulness of the features.



Note. For each feature, the left half (shaded blue) represents the participants with no prior experience with a feature. The right half (shaded orange) represents the ratings of those participants who either currently uses or previously used the feature.

5.4 Habituality of Media Use in Relation to Self-Control Tools

The SRBAI scores (see Table 2) were highest for social media and chatting, and lowest for gaming. A Shapiro-Wilk for the SRBAI values showed that the values were not normally distributed (p < 0.0001 for all SRBAI values). The number of participants for which this behaviour was not at all present (i.e., a score of 1) was highest for gaming (n = 102).

Table 2

Descriptive statistics

Behaviour	M	SD
SRBAI Gaming	7.27	3.96
SRBAI Chatting	11.89	4.68
SRBAI Social media	11.59	4.82
SRBAI Videos	9.63	4.38
SRBAI Surfing	8.83	4.37
MaxSRBAI	14.57	3.69

Table 3

Spearman Correlations between MaxSRBAI and the other measures

Feature	r	99% CI
Autoclose helpfulness	-.10	[-.36, .18]
Blocking helpfulness	-.09	[-.38, .22]
Gamification helpfulness	-.18	[-.49, .17]
Goals helpfulness	.05	[-.30, .39]
Pomodoro helpfulness	-.03	[-.24, .18]
Screenshare helpfulness	.28	[.06, .47]
Visualization helpfulness	-.37	[-.55, -.16]
Learning Satisfaction	-.08	[-.19, .04]
Self-Efficacy	-.22	[-.33, -.11]
Distraction Suffering	.45	[.35, .54]
Distraction Frequency	.49	[.39, .58]

Note. For the correlations with the DSCT features, only participants with prior experience with using a feature were considered. The features Delay, Modification and Compare are left out in this table due to their very low number of users with prior experience.

We correlated the perceived helpfulness of DSCT features with the MaxSRBAI scores (see Table 3). Only participants with prior experience with a feature were included in this analysis, as we aimed to investigate the actual experience rather than hypothetical perceptions. This analysis has an exploratory approach, and we interpret the confidence intervals of the correlations in the following manner: if a confidence interval does not include the zero, we interpret this as a sign that further research should investigate a particular correlation. Conversely, if a confidence interval includes 0, we interpret this cautiously as a lack of evidence supporting the effect.

For MaxSRBAI, we found a small negative correlation with Visualization, and a small positive correlation with Screenshare (see Table 3). All other correlations were small and included the zero.

5.5 Reasons for Stopping the Use of Self-Control Tools

The participants often stopped using a DSCT after they had used it for some time (see Table 1 in the “Quit” column),and elaborated on the reasons for this in free text responses. We have listed the codes for each tool in detail in the supplementary table under https://tinyurl.com/uba426xf. When we provide quotes of participant responses, we provide the id of the response signified by a hashtag. (e.g., #42 to denote that this was the response with the id 42).

5.5.1 Tools no longer needed

Stopping the use of a DSCT did not always indicate that the tool was not helpful for participants. The most common reason for stopping the use of a DSCT (n = 130) was that participants no longer needed it, either because they were in a different phase of their life or only required it temporarily. For instance, some participants reported that they only needed the DSCT during particularly stressful phases, as one participant noted, *"Because exams ended"* (#249). Other participants stopped using DSCTs because they felt that their self-control had improved sufficiently *("I felt that my social media usage decreased and I am not as distracted anymore"*, #164). However, we could not determine whether the improvement was due to using a DSCT, as participants did not provide enough information to make such conclusions.

A smaller group of participants (n = 15) stopped using DSCTs because they felt that their self-control had improved. These participants reported that they no longer needed the tools as they felt that they had gained greater control over their digital habits. However, it is unclear whether the participants' improvements were solely due to using the tools.

5.5.2 Lack of Motivation or Self-Control

Twelve responses indicated a self-reported lack of discipline, motivation, or self-control to DSCTs. These responses were most common for the Blocking and Autoclose tools, with one participant stating, *"I got bored and needed stimulation" (#72)*.

5.5.3 Balance of Restrictiveness

We also identified issues related to the balance of restrictiveness among participants' use of DSCTs. Specifically, 17 responses indicated that participants began to circumvent the restriction mechanisms put in place by the tools. For example, one participant stated,

"[...] I would find other ways to open distracting apps, e.g., opening the app on my tablet […]" (#134), while another participant mentioned that they *"[…] used to just extend the limit time [...]*" (#83).

Conversely, in 15 responses, participants indicated that they stopped using a tool because it was too restrictive and prevented access to content, they needed to complete their tasks, or to content they wanted to access during leisure time. For instance, one participant reported that the tool would "*block apps I needed when studying, e.g., YouTube*" (#114), while another participant said, *"[...] I use my phone to listen to music whilst I work and the app interferes with that*" (#112).

5.5.4 Negative Emotions in Tool Use

Our findings suggest that several participants experienced negative emotions when using their tools. Feelings of stress and/or anxiety when using DSCTs were mentioned in 15. Frequent self-reflection did not appear to be beneficial for everyone, as one participant stated, *"[Visualization] made me feel guilty for spending so much time on my phone [...]"* (#177).

Additionally, a small number of participants (n = 7) became annoyed by DSCTs, with the tools themselves becoming more of a distraction than an assistance. For example, one participant noted that a blocking tool "*distracted and restricted me more than it helped...*" (#108).

5.5.5 Feature-specific reasons

Further reasons where specific to a particular feature type. Gamification features lost their appeal for the participants in eight cases because the rewards ceased to be interesting for the participants. One participant noted, "*Because I had reached the maximum rewards and then I would have had to start all over again.*" (#299, translated). For DSCTs with goal reminders, the participants indicated four times that the notifications were too frequent or at the wrong time (*"Because at some point I was annoyed by the notifications*", #203). Furthermore, in four responses they also found that managing and keeping up with the goals became too much work. In the responses to Screenshare features, the participants noted that they had issues with their learning group. A recurring topic (n = 11) was that the group itself became a distraction ("*The meetings were rather complicated and not always productive.*", #164). In 13 responses, the participants mentioned that they had trouble finding a group for screensharing. For Pomodoro, 17 of the participants responded that using fixed time intervals did not fit their mode of working ("*It wasn't suited for the type of assignments I do. It breaks the flow of my thinking and isn't suited for time-consuming assignments like programming.*", #262).

6 Discussion

Our study found that the majority of participants in our sample experienced digital distractions that they perceived as interfering with the achievement of their learning goals. Self-control tools, which are promoted as a solution to this problem, are not universally perceived as helpful and students lack awareness of potentially helpful features. Our qualitative analysis of people's reasons for stopping DSCT use provides insights into why DSCTs fail for some users.

6.1 Use and Knowledge

In our sample, even the most popular DSCT features were far from universally known. Crucially, a small group of 8% of the students who suffered from distractions was not aware of any DSCT feature at all. Making DSCTs better known to this population could help more people. This is underlined by the observation that participants rated all features as more helpful when they had previous experience with them. Therefore, we believe that increased promotion of existing DSCTs is a necessary step to address the problem on a larger scale. Campaigns to increase awareness of the potential of DSCTs could be a good way to address this issue (I. Kim et al., 2017). However, there is the limitation that not all features are equally available on all devices. Especially Apple is highly restrictive in terms of apps allowed in the app store. For example, reading the usage times for other apps or preventing apps from opening in any way is prohibited for third-party apps. In relation to this situation, Lyngs and colleagues (2022) have noted that restrictive platforms must therefore take on the challenge of covering the multitude of possible user scenarios, as they cannot rely on third-party developers to do so.

6.2 Helpfulness

Our study revealed differences in the perceived helpfulness of different DSCT features, with the greatest discrepancy between the most helpful feature (Modification) and the least helpful feature (Visualization). Somewhat unexpected was the widespread use and positive reception of screen sharing. Despite the challenges of finding a good group to screen share with, this seems a promising strategy that deserves further attention. Overall, however, we caution against interpreting these results as an overall ranking, as each feature was found to be helpful by some of its users. Instead, we emphasize the importance of tailoring DSCTs to the specific needs and preferences of the user.

The results for the correlation between habitual usage and the perceived helpfulness of DCST features are somewhat in line with our expectations. Our findings did support the idea that pure self-awareness interventions, such as usage visualizations, are less helpful for highly habitual media use. The small positive correlation with Screenshare fits into the narrative as well, considering that engaging in an online session together with peers constitutes a change in the social context, which is well-suited to disrupting unwanted habits (Lally & Gardner, 2013).

On the other hand, we initially assumed that goal support features would be less helpful in mitigating habitual behaviour because these tools rely on individuals consciously comparing their actions with their goals, a process that is unlikely to occur in habitual behaviour (Fiorella, 2020). This would make such tools unsuitable for their purpose. However, this was not the case in our analyses. We suspect that one reason for the lack of negative correlation may be found in the specific tools that participants used. The specific tools that the participants named included many to-do lists and calendars that were not necessarily related to reducing digital distractions. We suspect that our feature description might not have been explicit enough, leading participants to report their use of general productivity tools.

6.3 Challenge of Dual-Purpose Platforms

A complication of DSCTs specific for use during learning is the dual purpose of platforms such as YouTube, which contain both educational and entertainment content. Using social media for educational purposes is well-known (Hrastinski & Aghaee, 2012), and our results show that this also has an influence on DSCT acceptance. Several participants stated that they required access to a certain platform, which was prevented by their blocker, and so they had to disable it. This makes smooth usage difficult to achieve and naturally leads to frustration.

A way to make platforms less distracting while keeping them usable would be the use of feature removal tools (e.g., the "HabitLab" tools by Kovacs et al. (2018)) that remove particularly distracting parts. Specifically for YouTube, which is certainly the platform that is currently most relevant as a dual-purpose platform, a recent study by Lukof et al. (Lukof et al., 2023) investigated the use of an adaptable commitment device. The users would start their visit on YouTube by explicitly stating their intent (e.g., entertainment or focus) and thus received a different interface. This was well received and led to greater satisfaction and goal alignment when using the platform. However, while this worked great in a study setting, it is doubtful that this will quickly find its way into practice, especially as a mobile app, as it is seemingly detrimental for the business case of platforms. Mobile apps are the greater challenge in this regard, because feature modifications do work quite well in browser extensions, as they simply have to modify the HTML content, which is comparatively trivial to do. The content in smartphone apps cannot be modified as easily for everyday use (Datta et al., 2022). Currently, users can only resort to installing alternative apps (Zhang et al., 2022), and these alternatives are often not available. Thus, although there are in principle good solutions for dual purpose platforms, in practice they are not available to all users.

6.4 The Various Reasons for Stopping DSCT Use

Overall, the reasons for discontinuing DSCTs supported the idea that the same tool does not work for everyone. For some participants who are unhappy with their tools, a reasonable first suggestion would be to try another tool to see if it suits them better. Especially as some pointed out quite trivial usability issues, e.g., Pomodoro timers that appear to be inflexible, where alternatives with more flexibility certainly exist. Again, this highlights that users should be made more aware of the large variety of existing DSCTs.

6.4.1 Seasonal and Infrequent Use

By far the most common reason why our participants stopped using their DSCTs was that they only needed them during stressful periods like exam preparation. This finding is in accordance with previous observations that DSCTs are often used to help with specific, undesirable, tasks rather than to limit usage time in general (Lyngs et al., 2022). With short-term and interval use of DSCTs, there is less chance of achieving lasting behaviour change. Shorter time frames are usually insufficient to break undesired habits or build beneficial ones, as habit formation typically require several weeks of context dependent cue-response repetitions (Lally et al., 2010; Stojanovic et al., 2020; Wood & Rünger, 2016). Therefore, users should not expect longer lasting behaviour change from this type of usage pattern. In such scenarios, user education is pivotal to understand the circumstances under which they can anticipate lasting behavioral transformation and when to expect only transient support.

6.4.2 The Role of Motivation in DSCT Use

The notion that DCST use in education is infrequent or seasonal also has implications for the understanding of motivation to use tools. Our observations point to the fact that a lack of motivation can lead to discontinuing DSCT use. Participants reported instances where they were would simply "lost motivation" about their tools, or instances where they got "too lazy" to continue using their DSCT. The motivation to change behavior also fluctuates in other domains, e.g., smoking (Zhou et al., 2009) or weight loss (Elfhag & Rössner, 2005). However, in contrast to these examples, users typically don't seek to permanently eliminate digital media from their lives (Lukoff et al., 2018; Monge Roffarello & De Russis, 2019), and the motivation to use a tool might only be temporary. This distinction necessitates a unique model of motivation for DSCT use, one that incorporates this temporal requirement.

Various theories and models may provide inspiration for constructing this motivational model, including the transtheoretical model of behavior change (Heller et al., 2013) or self-determination theory (Deci & Ryan, 2012). However, these models should be critically assessed for their applicability given the specific nature of the use of DSCT. For example, the transtheoretical model emphasizes the need for users to be willing to change. However, in the context of DSCT use, this willingness may fluctuate, peaking during periods of high pressure and waning during more relaxed periods. These unique, often transient, demands on DSCT need to be recognized in future research. Understanding users' motivations and how they change over time is crucial, particularly for adherence to DSCT use.

6.4.3 Managing DSCT Activation

The qualitative statements also show that it is a recurring problem that restrictive DSCTs require a certain amount of effort, and have to be repeatedly switched on and off. In the context of learning, this means that the tool has to be activated for the learning phases and then deactivated again for leisure media use. Over short periods of time, this leads to frustration and annoyance, and over longer periods of time, users may simply forget about their tools if they have not used them for a longer time.

A possible solution could be DSCTs that activate automatically. Context-aware (Schilit et al., 1994) activation has previously been explored in some instances. It has been promising for highly controlled workplace scenarios (Tseng et al., 2019), which are, however, not comparable to self-directed learning at home. Other context-aware triggers that have been under study were based on the time of day (Löchtefeld et al., 2013), the physical presence in a classroom (I. Kim et al., 2017) or the recognition of longer periods without movement (I. Kim et al., 2018). However, the context-awareness was only evaluated superficially, and we have doubts that these are sufficient and reliable triggers for the realities of everyday learning, where time and location of learning and entertainment vary and are interchangeable. If the triggers interfere too much with the learners and just get in their way, they will not find long lasting acceptance.

A potential improvement could be context-awareness of learning activities (Ciordas-Hertel et al., 2022). Given that learning materials are increasingly available on online learning platforms, a learner's activity could also serve as a contextual trigger to activate restrictions. A monitoring application could detect that a learning platform was visited, thereby starting the contextual activation. This could reduce the burden on users to frequently activate and deactivate their tools manually.

To mitigate the difficulties with dual-use platforms, adaptive DSCTs could furthermore experiment with content classification to distinguish between content that is relevant for learning and content that is purely for leisure. Machine Learning models for content classification have shown promising results in other domains (Yousaf & Nawaz, 2022) and could present a way to distinguish distractions from learning content without forcing the user to manage long lists of black- or whitelisted content.

6.5 Limitations and Suggestions for Future Research

A limitation to the generalizability of our results is that our sample of participants was limited to English and German speakers**.** Cultural differences in device use could yield different results in other regions (Gray & Schofield, 2021; Kononova & Chiang, 2015).

Some observations hint at the issue that our methodology for measuring helpfulness was insufficient to detect complex nuances like "This would help me in principle, however…". Another issue is that what feels good might not actually help to change behaviour. For example, users prefer non-restrictive interventions even when they are aware of their shortcomings (Zimmermann, 2021). The perceived helpfulness of an intervention may be high, and it may feel helpful, but not necessarily result in behaviour change (Patel et al., 2015).

While we cannot answer these questions regarding the effectiveness, our analysis of the reasons for stopping tool use yielded additional insights about DSCTs.

Another limitation is that combinations of DSCT features are very typical (Lyngs et al., 2022). For example, the “Digital Wellbeing” app on Android systems has usage visualizations, usage goals, and automatic closing features. We saw several such apps, like the “Forest” app, in multiple feature categories. We tried to circumvent this limitation by asking about features instead of specific tools, but disentangling the helpfulness of a single feature from a tool with multiple features can be challenging. Some participants stated that they stopped using a feature because it did not help them, but still rated it as helpful, which suggests that some of the participants could not sufficiently express their thoughts about the DSCTs.

A general limitation of cross-sectional studies on the topic of DSCTs is that they cannot incorporate the dynamics of the tool use. The participants often used DSCTs only for short, stressful phases, e.g., during exams. Afterward, they stopped again. This is understandable, but it potentially impacts the helpfulness of DSCTs because the tool usage does not become habitual. These phenomena can only be addressed with longitudinal studies that observe how the use of DSCTs waxes and wanes over a learner's career.

Moreover, our study focused on the perceived helpfulness and not the actual effectiveness. Surveys and self-reports are unlikely to be sufficient to measure the true effectiveness of DSCT features (Parry et al., 2021). Tracking digital activity requires a sophisticated tracking system. It must be able to track multiple devices simultaneously, otherwise, activity on the smartphone may be detected, but not activity on the notebook, or vice versa. It must also distinguish between cases where media use takes place during learning and those where it takes place during leisure (Authors, in press). In addition, as already mentioned, the role of internal states must be adequately considered, which requires frequent prompting of these states, for example via experience sampling (van Berkel et al., 2017). Such a tracking setup is technically challenging, but necessary to advance research on DSCTs in academic learning.

7 Conclusion

Digital self-control tools present a wide range of approaches to the urgent and growing problem of digital distractions. Our findings underscore that digital self-control tools can provide valuable assistance in mitigating digital distractions. However, their success is not universal; the perceived helpfulness varies significantly among individuals. One key obstacle impeding the utility of digital self-control tools is the lack of user awareness about available tools. This is especially noteworthy as there can be a mismatch between a user's needs and the features provided by a specific tool. Hence, it is essential for users to explore and find digital self-control tools that best align with their specific requirements.

Our qualitative assessment exploring why participants discontinued their use of these tools identified several general factors, such as the fine line between restrictiveness and accessibility, and the challenges posed by dual-purpose platforms. We also unveiled factors particular to certain tool features that can directly inform their enhancement.

In summary, while the use of self-control tools is widespread, numerous minor obstacles hinder their potential effectiveness in learning environments. Addressing these issues will substantially enhance the support we can provide to students dealing with digital distractions. As we navigate the expanding landscape of digital education, these considerations are key to leveraging information technologies to facilitate a more productive and focused learning experience.

Declarations

Ethical Approval

Ethics approval was obtained from the ethics committee of the DIPF | Leibniz-Institute for Research and Information in Education.

Informed Consent

All participants gave consent prior to participating in the study.

Data Availability

The survey data, with identifying data removed, is available via https://osf.io/8zmb7/?view_only=10162da5b5a54e18bd9a7e9830fd0638

Funding

No external funding was received to assist with the preparation of this manuscript.

Acknowledgement

Not applicable

References

Biedermann, D., Schneider, J., & Drachsler, H. (2021). Digital self-control interventions for distracting media multitasking - A systematic review. *Journal of Computer Assisted Learning*, *37*(5), 1217–1231. https://doi.org/10.1111/jcal.12581

Calderwood, C., Ackerman, P. L., & Conklin, E. M. (2014). What else do college students “do” while studying? An investigation of multitasking. *Computers & Education*, *75*, 19–29. https://doi.org/10.1016/j.compedu.2014.02.004

Chang, B. P. I., Webb, T. L., & Benn, Y. (2017). Why Do People Act Like the Proverbial Ostrich? Investigating the Reasons That People Provide for Not Monitoring Their Goal Progress. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00152

Ciordas-Hertel, G.-P., Biedermann, D., Winter, M., Mordel, J., & Drachsler, H. (2022). How can Interaction Data be Contextualized with Mobile Sensing to Enhance Learning Engagement Assessment in Distance Learning? *INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION*, 105–112. https://doi.org/10.1145/3536220.3558037

Datta, S., Kollnig, K., & Shadbolt, N. (2022). Mind-proofing Your Phone: Navigating the Digital Minefield with GreaseTerminator. *27th International Conference on Intelligent User Interfaces*, 523–536. https://doi.org/10.1145/3490099.3511152

Deci, E. L., & Ryan, R. M. (2012). *Motivation, Personality, and Development Within Embedded Social Contexts: An Overview of Self-Determination Theory*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195399820.013.0006

Elfhag, K., & Rössner, S. (2005). Who succeeds in maintaining weight loss? A conceptual review of factors associated with weight loss maintenance and weight regain. *Obesity Reviews*, *6*(1), 67–85. https://doi.org/10.1111/j.1467-789X.2005.00170.x

Fiorella, L. (2020). The Science of Habit and Its Implications for Student Learning and Well-being. *Educational Psychology Review*. https://doi.org/10.1007/s10648-020-09525-1

Gardner, B., Abraham, C., Lally, P., & de Bruijn, G.-J. (2012). Towards parsimony in habit measurement: Testing the convergent and predictive validity of an automaticity subscale of the Self-Report Habit Index. *International Journal of Behavioral Nutrition and Physical Activity*, *9*(1), 102. https://doi.org/10.1186/1479-5868-9-102

Granić, A., & Marangunić, N. (2019). Technology acceptance model in educational context: A systematic literature review. *British Journal of Educational Technology*, *50*(5), 2572–2593. https://doi.org/10.1111/bjet.12864

Gray, J., & Schofield, D. (2021). Media Multitasking: A Cross-Cultural Study. *International Journal of Computer Trends and Technology*, *69*(3), 64–73. https://doi.org/10.14445/22312803/IJCTT-V69I3P112

Heller, L. J., Skinner, C. S., Tomiyama, A. J., Epel, E. S., Hall, P. A., Allan, J., LaCaille, L., Randall, A. K., Bodenmann, G., Li-Tsang, C. W. P., Sinclair, K., Creek, J., Baumann, L. C., Karel, A., Andersson, G., Hanewinkel, R., Morgenstern, M., Puska, P., Bucks, R. S., … Denollet, J. (2013). Transtheoretical Model of Behavior Change. In M. D. Gellman & J. R. Turner (Eds.), *Encyclopedia of Behavioral Medicine* (pp. 1997–2000). Springer New York. https://doi.org/10.1007/978-1-4419-1005-9_70

Holte, A. J., & Ferraro, F. R. (2020). True colors: Grayscale setting reduces screen time in college students. *The Social Science Journal*, 1–17. https://doi.org/10.1080/03623319.2020.1737461

Hrastinski, S., & Aghaee, N. M. (2012). How are campus students using social media to support their studies? An explorative interview study. *Education and Information Technologies*, *17*(4), 451–464. https://doi.org/10.1007/s10639-011-9169-5

Jamet, E., Gonthier, C., Cojean, S., Colliot, T., & Erhel, S. (2020). Does multitasking in the classroom affect learning outcomes? A naturalistic study. *Computers in Human Behavior*, *106*, 106264. https://doi.org/10.1016/j.chb.2020.106264

Kim, I., Jung, G., Jung, H., Ko, M., & Lee, U. (2017). Let's FOCUS: Mitigating Mobile Phone Use in College Classrooms. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *1*(3), 1–29. https://doi.org/10.1145/3130928

Kim, I., Lee, U., & Cha, N. (2018). Exploring Context-aware Proactive Blocking for Distraction Management. *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers - UbiComp '18*, 1202–1207. https://doi.org/10.1145/3267305.3274108

Kim, J., Cho, C., & Lee, U. (2017). Technology Supported Behavior Restriction for Mitigating Self-Interruptions in Multi-device Environments. *Proceedings of the*

ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, *1*(3), 64:1-64:21. https://doi.org/10.1145/3130932

Kim, J., Jung, H., Ko, M., & Lee, U. (2019). GoalKeeper: Exploring Interaction Lockout Mechanisms for Regulating Smartphone Use. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *3*(1), 16:1-16:29. https://doi.org/10.1145/3314403

Kim, J., Park, J., Lee, H., Ko, M., & Lee, U. (2019). LocknType: Lockout Task Intervention for Discouraging Smartphone App Use. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, 1–12. https://doi.org/10.1145/3290605.3300927

Kim, Y.-H., Jeon, J. H., Choe, E. K., Lee, B., Kim, K., & Seo, J. (2016). TimeAware: Leveraging Framing Effects to Enhance Personal Productivity. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 272–283. https://doi.org/10.1145/2858036.2858428

Ko, M., Yang, S., Lee, J., Heizmann, C., Jeong, J., Lee, U., Shin, D., Yatani, K., Song, J., & Chung, K.-M. (2015). NUGU: A Group-based Intervention App for Improving Self-Regulation of Limiting Smartphone Use. *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 1235–1245. https://doi.org/10.1145/2675133.2675244

Kononova, A., & Chiang, Y.-H. (2015). Why do we multitask with media? Predictors of media multitasking among Internet users in the United States and Taiwan. *Computers in Human Behavior*, *50*, 31–41. https://doi.org/10.1016/j.chb.2015.03.052

Kovacs, G., Wu, Z., & Bernstein, M. S. (2018). Rotating Online Behavior Change Interventions Increases Effectiveness But Also Increases Attrition. *Proceedings of the ACM on Human-Computer Interaction*, *2*(CSCW), 1–25. https://doi.org/10.1145/3274364

Lally, P., & Gardner, B. (2013). Promoting habit formation. *Health Psychology Review*, *7*(sup1), S137–S158. https://doi.org/10.1080/17437199.2011.603640

Lally, P., Jaarsveld, C. H. M. van, Potts, H. W. W., & Wardle, J. (2010). How are habits formed: Modelling habit formation in the real world. *European Journal of Social Psychology*, *40*(6), 998–1009. https://doi.org/10.1002/ejsp.674

Löchtefeld, M., Böhmer, M., & Ganev, L. (2013). AppDetox: Helping users with mobile app addiction. *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia - MUM '13*, 1–2. https://doi.org/10.1145/2541831.2541870

Loid, K., Täht, K., & Rozgonjuk, D. (2020). Do pop-up notifications regarding smartphone use decrease screen time, phone checking behavior, and self-reported problematic smartphone use? Evidence from a two-month experimental study. *Computers in Human Behavior*, *102*, 22–30. https://doi.org/10.1016/j.chb.2019.08.007

Lukof, K., Lyngs, U., Shirokova, K., Rao, R., Tian, L., Zade, H., Munson, S., & Hiniker, A. (2023). *SwitchTube: A Proof-of-Concept System Introducing "Adaptable Commitment Interfaces" as a Tool for Digital Wellbeing*. https://doi.org/10.1145/3544548.3580703

Lukoff, K., Yu, C., Kientz, J., & Hiniker, A. (2018). What Makes Smartphone Use Meaningful or Meaningless? *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *2*(1), 22:1-22:26. https://doi.org/10.1145/3191754

Lyngs, U., Lukoff, K., Csuka, L., Slovák, P., Van Kleek, M., & Shadbolt, N. (2022). The Goldilocks level of support: Using user reviews, ratings, and installation numbers to investigate digital self-control tools. *International Journal of Human-Computer Studies*, *166*, 102869. https://doi.org/10.1016/j.ijhcs.2022.102869

Lyngs, U., Lukoff, K., Slovak, P., Binns, R., Slack, A., Inzlicht, M., Van Kleek, M., & Shadbolt, N. (2019). Self-Control in Cyberspace: Applying Dual Systems Theory to a Review of Digital Self-Control Tools. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–18. https://doi.org/10.1145/3290605.3300361

Lyngs, U., Lukoff, K., Slovak, P., Seymour, W., Webb, H., Jirotka, M., Zhao, J., Van Kleek, M., & Shadbolt, N. (2020). "I Just Want to Hack Myself to Not Get Distracted": Evaluating Design Interventions for Self-Control on Facebook. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–15. https://doi.org/10.1145/3313831.3376672

Masood, A., Luqman, A., Feng, Y., & Ali, A. (2020). Adverse consequences of excessive social networking site use on academic performance: Explaining underlying mechanism from stress perspective. *Computers in Human Behavior*, *113*, 106476. https://doi.org/10.1016/j.chb.2020.106476

Monge Roffarello, A., & De Russis, L. (2019). The Race Towards Digital Wellbeing: Issues and Opportunities. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/3290605.3300616

Nguyen, V. T. (2022). The perceptions of social media users of digital detox apps considering personality traits. *Education and Information Technologies*, *27*(7), 9293–9316. https://doi.org/10.1007/s10639-022-11022-7

Nikken, P., & De Haan, J. (2015). Guiding young children's internet use at home: Problems that parents experience in their parental mediation and the need for parenting support. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, *9*(1). https://doi.org/10.5817/CP2015-1-3

Parry, D. A., Davidson, B. I., Sewall, C. J. R., Fisher, J. T., Mieczkowski, H., & Quintana, D. S. (2021). A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. *Nature Human Behaviour*, *5*(11), 1535–1547. https://doi.org/10.1038/s41562-021-01117-5

Pinder, C., Vermeulen, J., Cowan, B. R., & Beale, R. (2018). Digital Behaviour Change Interventions to Break and Form Habits. *Acm Transactions on Computer-Human Interaction*, *25*(3), 15. https://doi.org/10.1145/3196830

Ratan, Z. A., Parrish, A.-M., Zaman, S. B., Alotaibi, M. S., & Hosseinzadeh, H. (2021). Smartphone Addiction and Associated Health Outcomes in Adult Populations: A Systematic Review. *International Journal of Environmental Research and Public Health*, *18*(22), 12257. https://doi.org/10.3390/ijerph182212257

Roffarello, A. M., & De Russis, L. (2022). Achieving Digital Wellbeing Through Digital Self-Control Tools: A Systematic Review and Meta-Analysis. *ACM Transactions on Computer-Human Interaction*. https://doi.org/10.1145/3571810

Rosen, L. D., Mark Carrier, L., & Cheever, N. A. (2013). Facebook and texting made me do it: Media-induced task-switching while studying. *Computers in Human Behavior*, *29*(3), 948–958. https://doi.org/10.1016/j.chb.2012.12.001

Rubinstein, J. S., Meyer, D. E., & Evans, J. E. (2001). Executive control of cognitive processes in task switching. *Journal of Experimental Psychology. Human Perception and Performance*, *27*(4), 763–797. https://doi.org/10.1037//0096-1523.27.4.763

Schilit, B., Adams, N., & Want, R. (1994). Context-Aware Computing Applications. *1994 First Workshop on Mobile Computing Systems and Applications*, 85–90. https://doi.org/10.1109/WMCSA.1994.16

Stojanovic, M., Grund, A., & Fries, S. (2020). App-Based Habit Building Reduces Motivational Impairments During Studying – An Event Sampling Study. *Frontiers in Psychology*, *11*, 167. https://doi.org/10.3389/fpsyg.2020.00167

Tandon, P. S., Zhou, C., Hogan, C. M., & Christakis, D. A. (2020). Cell Phone Use Policies in US Middle and High Schools. *JAMA Network Open*, *3*(5), e205183. https://doi.org/10.1001/jamanetworkopen.2020.5183

Terry, C. A., Mishra, P., & Roseth, C. J. (2016). Preference for multitasking, technological dependency, student metacognition, & pervasive technology use: An experimental intervention. *Computers in Human Behavior*, *65*, 241–251. https://doi.org/10.1016/j.chb.2016.08.009

Tran, J. A., Yang, K. S., Davis, K., & Hiniker, A. (2019). Modeling the Engagement-Disengagement Cycle of Compulsive Phone Use. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/3290605.3300542

Tseng, V. W.-S., Lee, M. L., Denoue, L., & Avrahami, D. (2019). Overcoming Distractions during Transitions from Break to Work using a Conversational Website-Blocking System. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, 1–13. https://doi.org/10.1145/3290605.3300697

van Berkel, N., Ferreira, D., & Kostakos, V. (2017). The Experience Sampling Method on Mobile Devices. *ACM Computing Surveys*, *50*(6), 93:1-93:40. https://doi.org/10.1145/3123988

Verplanken, B., & Orbell, S. (2003). Reflections on past behavior: A self-report index of habit strength. *Journal of Applied Social Psychology*, *33*, 1313–1330. https://doi.org/10.1111/j.1559-1816.2003.tb01951.x

Wammes, J. D., Ralph, B. C. W., Mills, C., Bosch, N., Duncan, T. L., & Smilek, D. (2019). Disengagement during lectures: Media multitasking and mind wandering in university classrooms. *Computers & Education*, *132*, 76–89. https://doi.org/10.1016/j.compedu.2018.12.007

Wood, W., & Rünger, D. (2016). Psychology of Habit. *Annual Review of Psychology*, *67*(1), 289–314. https://doi.org/10.1146/annurev-psych-122414-033417

Yousaf, K., & Nawaz, T. (2022). A Deep Learning-Based Approach for Inappropriate Content Detection and Classification of YouTube Videos. *IEEE Access*, *10*, 16283–16298. https://doi.org/10.1109/ACCESS.2022.3147519

Zhang, M. R., Lukoff, K., Rao, R., Baughan, A., & Hiniker, A. (2022). Monitoring Screen Time or Redesigning It?: Two Approaches to Supporting Intentional Social Media Use. *CHI Conference on Human Factors in Computing Systems*, 1–19. https://doi.org/10.1145/3491102.3517722

Zhou, X., Nonnemaker, J., Sherrill, B., Gilsenan, A. W., Coste, F., & West, R. (2009). Attempts to quit smoking and relapse: Factors associated with success or failure from the ATTEMPT cohort study. *Addictive Behaviors*, *34*(4), 365–373. https://doi.org/10.1016/j.addbeh.2008.11.013

Zimmermann, L. (2021). "Your Screen-Time App Is Keeping Track": Consumers Are Happy to Monitor but Unlikely to Reduce Smartphone Usage. *Journal of the Association for Consumer Research*, *6*(3), 377–382. https://doi.org/10.1086/714365

Paper 5

Biedermann, D.*, Breitwieser, J*., Nobbe, L., Drachsler, H., and Brod, G. (2023). Designing an app to enhance children's planning skills: A case for personalized technology. [This paper is in its second revision in the journal *Behavior and Information Technology*]

Designing an app to enhance children's planning skills:

A case for personalized technology

Daniel Biedermann*[1], Jasmin Breitwieser*[2], Lea Nobbe[2], Hendrik Drachsler[1,3] and Garvin Brod[2,4]

[1] Information Center for Education, DIPF | Leibniz Institute for Research and Information in Education, Frankfurt am Main, Germany

[2] Education and Human Development, DIPF | Leibniz Institute for Research and Information in Education, Frankfurt am Main, Germany

[3] Computer Science Faculty, Goethe University Frankfurt, Frankfurt am Main, Germany

[4] Institute of Psychology, Goethe University Frankfurt, Frankfurt am Main, Germany

Author Note

Daniel Biedermann https://orcid.org/0000-0001-9219-222X

Jasmin Breitwieser https://orcid.org/0000-0002-0337-7159

Lea Nobbe https://orcid.org/0000-0002-5400-1952

Hendrik Drachsler https://orcid.org/0000-0001-8407-5314

Garvin Brod https://orcid.org/0000-0002-7976-5609

Daniel Biedermann and Jasmin Breitwieser contributed equally to this work.

The study was preregistered on the Open Science Framework (https://osf.io/w6pdv?view_only=b3dba1b617d3408396f66badc7a236ef). The data and analyses scripts are available at https://osf.io/sbcfd/?view_only=cd294d0742a7422a99a460fd22541388.

We have no conflicts of interest to disclose.

Our research was supported by the Distr@l program, Hessian State Chancellery. G.B. was supported by a Jacobs Foundation Early Career Research Fellowship. We thank Solveig Gresselmeyer for her invaluable help in designing the app.

Correspondence concerning this article should be addressed to Daniel Biedermann, Information Center for Education, DIPF | Leibniz Institute for Research and Information in Education, Rostocker Str. 6, 60323 Frankfurt am Main, Germany. E-mail: d.biedermann@dipf.de

Abstract

Planning is an important but difficult self-regulation strategy. The successful implementation of a plan requires that the plan is retrievable in everyday life when it is needed. Children in particular are unlikely to use effective strategies to internalize plans in a way that makes them easy to remember. Therefore, we designed PROMPT, a planning app to help children create and internalize plans effectively. The app included different internalization activities that were hypothesized to promote deeper or shallower processing of plans. School-aged children (N = 106, 9-14 years) used PROMPT for 27 days in their daily lives. Contrary to our hypotheses, the type of internalization activity was not associated with memory success overall. Deeper processing activities were only effective for children who spent more time performing these activities, suggesting that there were differences in how effectively children could make use of the internalization activities. These individual differences were predicted by children's grade level and their analogical reasoning abilities and mediated by time on task. Findings suggest that a child-appropriate planning app needs to be personalized to be effective; internalization activities have to be tailored to children's learning prerequisites.

Keywords: planning; implementation intentions; internalization; individual differences; mobile application

1 Introduction

Against the backdrop of the virtually ubiquitous proliferation of digital devices, there are clear signs that this is increasingly posing challenges to children's self-regulatory abilities. Even before the Covid-19 pandemic and the increase in emergency remote teaching (Drachsler et al., 2021; Lozano-Blasco et al., 2022) there was evidence of an increase in excessive internet and smartphone use among children (Cho & Lee, 2017). These challenges indicate that there is a need to support children's self-regulation skills (Mahapatra, 2019).

However, mobile devices do not necessarily have to play a harmful role in this context, but could also provide help in the form of applications that support children's use of self-regulation strategies. An important self-regulation strategy that can be supported with digital applications is planning. Setting goals and planning are essential steps of self-regulated learning (Gollwitzer, 1990; Zimmerman, 2002) that have proven to be successful in combination with mobile devices in higher education settings (Tabuenca et al., 2015). However, creating effective plans, remembering them when needed and following through are not easy tasks. Studies have shown that mental contrasting with implementation intentions (MCII) is an effective intervention to help children achieve their academic and non-academic goals (Wang et al., 2021). MCII means that children commit to their goals by mentally contrasting the desired future with present reality (Oettingen et al., 2001) and make "simple plans" (Gollwitzer, 1999) that drive action implementation.

There are already apps that address how to support planning on apps. For instance, the WOOP app (https://woopmylife.org) guides users to perform MCII. However, the WOOP app is not specifically designed for children, nor are there studies testing its effectiveness in children. We argue that a child-appropriate planning app needs to place a specific emphasis on children's needs and requirements to deeply internalizing their plans because their prospective memory and metacognition are far below the level of adults (e.g., Kvavilashvili et al., 2008; Schneider, 2010). For example, using the WOOP app, children could create an implementation intention to turn off their smartphone as soon as they start studying to prevent unwanted distractions. To put this plan into action, they need to retrieve it as soon as they start studying. If they only recall their plan when their smartphone receives a notification, the plan has already failed at preventing the distraction. When designing a planning app for children, it is thus important to keep in mind that *creating* effective plans is only half the battle: the crux of success lies in the actual internalization and implementation of the plans. The smartphone and its rewarding apps that are just one gesture away pose a serious distraction especially for young children that have less self-control (Johannes et al., 2019; Ward et al., 2017).

1.1 Internalization as an Important Step of Planning

Planning is generally considered an important step of self-regulation that facilitates tasks requiring self-control (Gollwitzer, 1996). A plan can be conceptualized as a metacognitive knowledge structure stored in declarative memory that specifies when to do what (i.e., corresponding to a production rule in the ACT-R framework, Anderson et al., 2004, or the metacognitive WWW&H rule, Veenman et al., 2006). Implementation intentions, for instance, create an associative link between a situation and an action in memory (Gollwitzer, 1999). It is assumed that, once the situation is encountered, the behavior is initiated automatically on the basis of this associative link (Brandstätter et al., 2001; Webb & Sheeran, 2007). Thus, the mental representation of a plan in memory needs to be sufficiently strong so that the plan is retrieved and implemented at the right time. Studies have used different activities to support the internalization of a plan, such as asking people to re-read the plan (e.g., Breitwieser et al., 2021), to say it aloud (e.g., Cohen et al., 2007), or to copy it on a sheet of paper (e.g., Hoch et al., 2020). However, there is a lack of research systematically comparing these different internalization activities (Hagger et al., 2016), and it is unclear how to best strengthen the mental representation of a plan in memory.

What are the crucial cognitive processes for successful internalization of a plan? It has been argued that creating a mental representation of a plan requires attention and other cognitive capacities during the planning phase (Martiny-Huenger et al., 2016). Based on information processing theories, however, simply allocating cognitive resources to the processing of a plan is not sufficient to ensure a strong mental representation of it; the *level of processing* also matters (Craik & Lockhart, 1972). For instance, according to the levels of processing theory, reading and repeating a plan without elaboration should create a weaker mental representation of the plan than reflecting on its meaning and creating associations with prior knowledge. This assumption is supported by studies that combined implementation intentions with mental imagery (Fennis et al., 2011; Knäuper et al., 2009; Oh & Larose, 2015): Vividly imagining the situation and the behavior specified by the plan enhances its effectiveness compared to only re-reading the plan.

In real-world settings, people often try to bypass the step of creating a strong mental representation of a plan in their own memory. One common strategy is to use digital reminders as external aids to trigger the execution of a task (Brewer et al., 2017; Tabuenca et al., 2015). However, relying on reminders has several shortcomings: First, reminder notifications are often ignored or not noticed at all (Visuri et al., 2019). Second, time-based reminders need to be pre-scheduled, but the time at which the critical situation will arise may not be known in advance or may change too quickly. Third, event-based reminders (e.g., Pinder et al., 2016) fail if a plan tackles an internal state (e.g., a detrimental thought) rather than an external context (Achtziger et al., 2008; Breitwieser et al., 2021). Fourth, linking the goal-directed action to a reminder rather than a situational cue creates an undesirable dependency on the reminder (Stawarz et al., 2015; Wicaksono, Beale, et al., 2019). Finally, the smartphone itself poses a distraction to studying (Johannes et al., 2019; Ward et al., 2017). It would be absurd to send students a notification during studying that reminds them to ignore their phone. Taken together, relying on reminders is a suboptimal solution to help people recall their plans at the right moment.

Due to their shortcomings, reminders should not be used to trigger an intended action when it is supposed to happen, but to remind people of their plans (Pirolli et al., 2017; Wicaksono, Hendley, et al., 2019). Plan reminders are thus used to improve memory for a plan. Recent research suggests that plan reminders might need to be presented quite frequently (e.g., daily) to prevent the effect of a plan from diminishing (Breitwieser et al., 2021). Typical plan reminders come in the form of text notifications that encourage re-reading of the plan (e.g., Kramer et al., 2020; Wicaksono, Beale, et al., 2019). However, due to the lack of systematic research on this topic, the question remains whether this activity is really the most effective way to stimulate deep internalization of plans.

1.2 Stimulating Deep Internalization of Plans

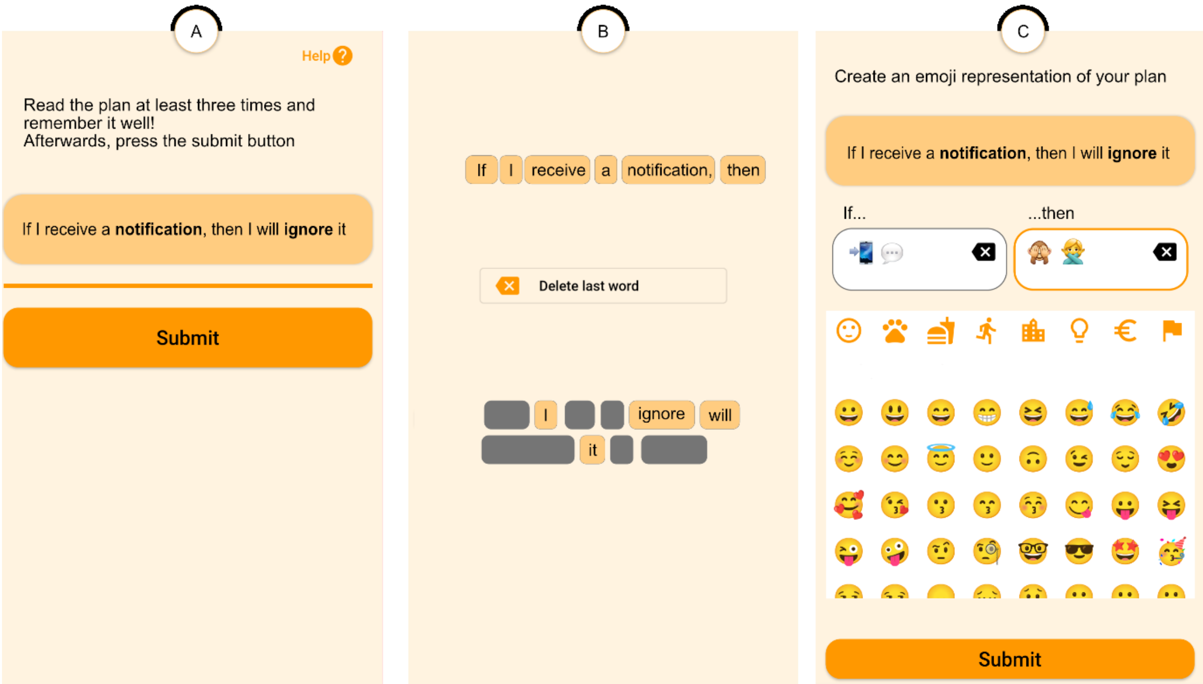
Which internalization activity could be implemented in an app to stimulate deep processing of plans? To answer this question, we draw on the literature on learning strategies. Specifically, the ICAP framework (Chi et al., 2018; Chi & Wylie, 2014) provides a useful taxonomy that categorizes learning strategies in terms of the cognitive processes that they involve. This taxonomy thus allows us to make predictions about which internalization activities may promote deeper or shallower processing of plans. To achieve deep levels of processing, the goal is to increase children's level of engagement with the to-be-learned material (here: the plans). In the present context, the often rather ill-defined term of "engagement" in human-computer interactions (Doherty & Doherty, 2019) thus takes on the specific meaning of *cognitive engagement* while interacting with a planning task.

The four modes of cognitive engagement that are distinguished within the ICAP framework are passive, active, constructive, and interactive. They form a hierarchy, where it is assumed that interactive engagement leads to deeper processing (and thus better learning outcomes) than constructive engagement, which in turn leads to deeper processing than active engagement, which leads to deeper processing than passive engagement (I > C > A > P). Passive engagement describes activities where information is received without overtly performing an action (e.g., reading), while active engagement involves the manipulation of the content (e.g., underlining). Constructive engagement further requires that the learner generates output that goes beyond the initial information (e.g., illustrations) and interactive engagement requires that at least two parties co-construct this output (e.g., via dialogue). Thus, building on models of processing depth (Craik & Lockhart, 1972), the ICAP framework provides a clear taxonomy that predicts which learning strategies correspond to which level of processing depth.

In the present study, we aimed to translate the ICAP taxonomy into activities for internalizing plans that can be implemented in a child-friendly app. However, we did not implement an interactive internalization activity because we wanted to first evaluate the effectiveness of constructive engagement before further developing an activity derived from it (i.e., co-construction). As outlined in the previous section, typical plan reminders stimulate passive engagement by prompting users to re-read their plan (Kramer et al., 2020; Wicaksono, Beale, et al., 2019). Therefore, we adopted this activity as our passive internalization activity ("reading activity"). Children were asked to re-read the plan without taking further action (Fig. 1a). To stimulate active engagement, we asked children to recall the plan immediately after reading it as a form of retrieval practice. To implement this internalization activity as an app-based intervention, we took inspiration from popular learning apps such as Duolingo, which ask users to reassemble a list of scrambled words in the correct order ("puzzle activity", Fig. 1b). This activity qualifies as active engagement because children had to actively manipulate the content. In contrast to typing a plan via the on-screen keyboard and then reading it (see, for instance, the WOOP app), this also avoids asking children to type on an on-screen keyboard, which is slow, has a high error rate and is often disliked as a means to input text (Smith & Chaparro, 2015). To stimulate constructive engagement and thus even deeper processing of the plans, we asked children to create a representation of their plan by using emojis ("emoji activity", Fig. 1c). This activity qualifies as constructive engagement because children had to generate content that goes beyond the presented information (i.e., the plan; see Brod, 2021). Based on the ICAP framework, we expected children to internalize and thus later remember the plans best with the emoji activity, second best with the puzzle activity, and least with the reading activity.

Figure 1

Screenshots of the Reading (A), Puzzle (B), and Emoji (C) Activity



1.3 Designing a Child-Appropriate Planning App

Children in particular may benefit from an easily accessible app that helps them make and internalize plans. Research suggests that prospective memory - the ability to remember and implement plans at the right moment in the future - still develops across childhood (Kvavilashvili et al., 2008). Furthermore, due to ongoing development of metacognitive abilities (Bjorklund et al., 2008; Schneider, 2010), children are unlikely to select and engage in effective planning strategies on their own.

However, designing a planning app for children that they can use in their everyday lives also poses additional challenges. That is, to have an impact in real-world settings, children must use the app voluntarily and effectively without the supervision of a third party. For the deeper internalization activities to be effective, it is important that children are willing to actually perform the actions and, even more importantly, engage in the cognitive processes the activities aim to stimulate (Chi & Wylie, 2014; Fiorella & Mayer, 2016). What factors may influence whether children use the app-based internalization activities voluntarily and effectively?

From a human-computer interaction perspective, the Technology Acceptance Model (TAM; Davis et al., 1989) provides a framework to investigate people's willingness to engage with new technology. According to the TAM, people's attitude toward a new technology is determined by two beliefs: the belief that using the technology is effortless (i.e., perceived ease of use), and the belief that using the technology will enhance performance (i.e., perceived usefulness). The resulting attitude is a person's affective reaction to the technology (e.g., enjoyment, see Venkatesh et al., 2003). As an extension of Ajzen and Fishbein's (1980) theory of reasoned action, the TAM assumes that attitude is the main predictor of the behavioral intention to use a technology, which in turn predicts actual use. Therefore, children's attitudes toward using the internalization activities are a critical component in evaluating the child-appropriateness of the activities.

From a developmental psychology perspective, the design of a child-appropriate planning app must align with children's individual characteristics. Importantly, a technological design that is appropriate for one child may not be appropriate for another due to motoric and (meta-)cognitive abilities that are still developing in children (Brod, 2021; Martens, 2012). In the present context, the different internalization activities implemented in the planning app do not only include different overt actions (e.g., unscrambling sentences vs. picking emojis) but also different *cognitive* operations (i.e., deeper or shallower processing of the plans). Research suggests that the effectiveness of these activities could vary greatly between children due to developmental differences in (meta-)cognitive prerequisites (for an overview, see Brod, 2021). For instance, the control of memory processes requires executive functions that are improving throughout childhood and adolescence (Diamond, 2013) and could thus lead to age-related differences in the effectiveness of retrieval strategies (e.g., Aslan & Bäuml, 2016; Breitwieser & Brod, 2021). Other strategies might have other cognitive prerequisites (Brod, 2021); still others might have motivational rather than cognitive prerequisites (Jacob et al., 2022). These examples illustrate that a child-appropriate app that includes learning activities to internalize plans may need to be tailored to children's individual prerequisites.

1.4 The Present Study

The goal of this study was to find out how a planning app should be designed to best help children internalize plans. Because successful implementation of a plan depends on its representation in memory (Hagger et al., 2016; Martiny-Huenger et al., 2016), we specifically focused on the effectiveness of different app-based internalization activities for promoting children's retrieval of plans. To test the effectiveness of these app-based activities, children used our planning app under realistic conditions in their everyday lives without the supervision of an experimenter. Children had to internalize a different plan each day for up to 27 days, switching between three internalization activities. Our primary outcome measure was recall performance, which was assessed at least six hours after internalizing a plan.

The study was pre-registered at the Open Science Framework (https://osf.io/w6pdv/?view_only=b3dba1b617d3408396f66badc7a236ef). Based on the ICAP framework, we hypothesized that children would recall the plans best when using an internalization activity that stimulates constructive rather than active engagement (H1a; emoji > puzzle), and when using an activity that stimulates active rather than passive engagement (H1b; puzzle > reading). We further explored children's attitude toward using the internalization activities.

In addition, we explored how much effort children actually invested when using the active and constructive internalization activities, using time on task as a proxy measure. Especially in the emoji condition, carefully performing the activity (i.e., selecting emojis that fit the meaning of the plan) should take more time than careless selection. In the puzzle condition, time on task indicates the difficulty of the activity since children could only continue when they had unscrambled the sentence correctly. In the next step, we explored whether children's prerequisites, specifically their grade level and cognitive abilities, predicted which internalization activity helped them most in recalling the plans. Cognitive abilities were measured in a follow-up assessment two months after the main study to further explore inter-individual differences in the effectiveness of the internalization activities. We focused on executive functions because they play an important role in memory processes and are late developing (Diamond, 2013). Since using emojis to represent the meaning of a sentence is a form of analogy use, we further assessed children's analogical reasoning abilities.

2 Methods

2.1 Participants

We recruited German-speaking children in grades 4 to 7 (late primary and early secondary level in the German school system) in Spring 2021. Children were recruited via the social media platforms of our institute, a newsletter that was sent to children using a vocabulary learning app[9], email distributors of parents' councils across Germany, and flyers distributed in the area of Mainz (Germany). Children's legal guardians were forwarded to an online form through which they gave informed consent and registered their children for participation; 183 children were registered, of whom 114 participated in the study and completed at least one assessment.

As stated in the preregistration, we only excluded children if their data was implausible or indicated that the tasks were not completed as intended. We excluded eight children who had an average recall performance of 0% and, when we reviewed their raw data, appeared not to even have tried to remember the plans correctly. One child's language level was reported as only "moderate" by its parent but this child's average recall performance (25%) was within 1 SD of the mean and we thus saw no reason to exclude it. We excluded five data points because the time span between internalization and recall of the plan was only 2 minutes (n = 1) or because children spent less than 5 s (i.e., less time than needed to read the plan at least once) using the internalization strategy (n = 4). Thus, the final sample for the analyses presented here consists of 106 children (58% female; M_{age} = 11.80 years, SD_{age} = 1.04 years) with a total of 2134 plan recalls (M = 20.13 recalls per child, SD = 7.20). Due to difficulties in recruitment, this sample size is smaller than the 138 participants we had aimed for based on our power analysis reported in the preregistration.

As compensation for their participation, children received a gift card for an online shop of their guardian's choice. The value of the gift card varied between 50 ct and 12 Euro

[9] For the present study, it was not important whether children used the vocabulary learning app or not. Deviating from the preregistration, we therefore also recruited children who did not use the vocabulary learning app to increase our sample size.

depending on a child's participation rate. Ethics approval was obtained from the ethics committee of DIPF | Leibniz Institute for Research and Information in Education.

2.2 Materials and Measures

This section describes the materials and measures used in the study.

2.2.1 Sociodemographic Information

Sociodemographic information was collected from children's legal guardians in an online questionnaire after they had filled out the consent form. The questionnaire asked for children's gender, birth date, school type, grade level, federal state, and German language level. Data of three children are missing. The birth dates of six children were not plausible because they reported ages less than 1 or greater than 40.

2.2.2 Usability Pre-Study

To rule out usability and comprehensibility issues with the app, we conducted usability tests with four children in the same age range as our target sample. A sample size range of three to five has been suggested as sufficient in uncovering usability problems for software interfaces (J. R. Lewis, 1994; Virzi, 1992). For the usability tests, we created a version of the app where all screens of the app could be opened individually instead of having to wait for a specific time like in the actual study. The screens included the onboarding process, the daily questionnaires, and the planning and recall activities that we describe in detail in the following sections. Furthermore, we presented all 27 plans to the children and asked whether they understood their meaning. An instructor was present for any questions or in case of technical problems, and we filmed the smartphone screens in order to be able to trace which inputs the children made at each step. We used a think-aloud protocol based on Lewis (C. Lewis, 1982) and, therefore, instructed the children to say out loud whatever they thought about the app and whenever they had any questions. Apart from that, we only provided the in-app instructions. Except for one child who was tested in person, sessions were conducted remotely with the webcam pointed at the smartphone screen. The studies did not show any usability problems with any of the activities and children reported feeling comfortable using them. Children did not report any difficulties in understanding the 27 plans.

2.2.3 Internalization Activities

The study was conducted in the PROMPT app that we developed as a research platform. Within the app, children had to learn pre-created plans that all followed the typical “if-then” format of an implementation intention (Gollwitzer, 1999). All plans had a similar sentence length and complexity (see supplement). The critical words of the if- and then-part of the plan were printed in boldface. Plans were presented within the app with one of three internalization activities: emoji, puzzle, or reading (see Figure 1).

In the emoji activity, children were asked to create a representation of their plan using emojis. The plan was displayed at the top of the screen, and the emojis had to be entered into separate input fields for the “if”-part of the plan and for the “then”-part of the plan. By tapping on the left ("if") input field, it would be highlighted, and the emojis the child selected would be inserted into that field. The same applied to the right ("then") input field. This allowed children to enter emojis for each part of the plan independently. To avoid the effects of idiosyncrasies in smartphone on-screen keyboards, we developed a custom emoji picker that contained emojis found in popular messaging apps. Typically, built-in emoji pickers on smartphones display recently and frequently used emojis more prominently, but we wanted to prevent this for our study. Overall, our emoji picker was structured as it is common in chat apps or similar: One could scroll through the emojis from top to bottom or jump more quickly to certain categories of emojis by selecting category icons at the top of the emoji picker. Children who are familiar with chat apps should be able to find the emojis they use in everyday life. The children could move on to the next screen after an emoji was entered into both the “if” and the “then” input fields.

In the puzzle activity, the plan was displayed for eight seconds before fading out. After the plan disappeared, it was segmented into individual words that were displayed in random order on the screen. By tapping on one of these words, it was added to the finished sentence. The children could only move on to the next screen when they had puzzled the complete sentence together correctly.

In the reading activity, the plan was displayed together with instructions to read it through at least three times. At the earliest, the children could move on to the next screen after 15 seconds had elapsed. The time was presented in the form of a loading bar.

On all screens within the app, we collected timestamps for the start and the end of the activity. We used the difference between starting and leaving an internalization activity to operationalize the time on task for each internalization activity. Because the average time on task varied greatly between the three internalization activities, outliers were determined for each activity separately. For each internalization activity, times were marked as outliers if they lay outside the 1.5 Tukey fences (Tukey, 1977). These outliers were excluded from all analyses involving time on task (pairwise exclusion; emoji: 5.18%, puzzle: 10.10%, reading: 15.34% of data points).

2.2.4 Recall Task

Recall performance for the plans was tested with a recall task that children were asked to perform six hours after the internalization of a plan. The app sent a notification after six hours to remind children of the task. The instruction on the screen was to write down the plan as they remembered it. Input fields were separate for the "if" and the "then" part of the plan to cue children and reduce the typing burden.

We used a two-step process to rate the responses, first an automatic one, followed by manual rating for those responses that we could not confidently rate automatically. For the automatic rating, we tokenized the sentences, removed stop words and punctuation and rated those as incorrect which had a remaining input length of only one word or less (n = 259). Next, we calculated an edit distance between the plan and the response, where we rated everything with an edit distance smaller than 3 as correct (n = 704). The remaining 1,402 responses were coded manually.

For the manual rating, two raters coded the remaining responses as correct (1) or incorrect (0) according to the following pre-registered criteria: Responses were coded as correct if a) they included the pre-specified cue (if-field) and target words (then-field) that were printed in boldface during the internalization activity (see Figure 1), b) the correct answer was paraphrased with slightly different words, but conveyed the same meaning. Responses were coded as incorrect if a) either cue or target word were missing or incorrect (i.e., not conveying the same meaning), b) cue and target word were assigned to the wrong input field (i.e., the word for the if-field was typed in the then-field or vice versa), c) cue and target word were correct but the remaining words invalidated the correct meaning (e.g., the addition of "not" or "no" when the correct sentence was positively worded). Responses that the raters had coded differently (339 sentences, 24.18%) were discussed by the first and third author until consensus was achieved.

2.2.5 Attitude

Children's attitude toward the internalization activities was assessed with three items each time children had used an activity three consecutive times to internalize the plans. That is, children who completed all 27 recalls reported their attitude nine times (three times per internalization activity) throughout the study. The items captured children's enjoyment ("Did you enjoy learning the plans with this activity?"), perceived usefulness ("Were you able to remember the plans better with this activity?"), and perceived effort ("How exhausting was it to remember the plans with this activity?"). Children answered each item on a 5-point Likert scale (1 = "not at all", 5 = "very"). We recoded the item on perceived effort. The mean scores of the three items correlated strongly (range: .61 - .67). We calculated the mean of the three items to obtain a score for children's attitude toward the internalization activities.

Furthermore, at the end of the study, we asked children which internalization activity or combination of activities they would prefer to use if they had to learn the plans again every day (forced choice). Children could either choose one of the individual internalization activities (emoji, puzzle, or reading) or they could choose to alternate between two or all three activities (7 response options in total). Ninety-one out of 106 children indicated their preference.

2.2.6 Cognitive Abilities

Cognitive abilities were assessed two months after the end of the 27-day ambulatory assessment period to explore cognitive explanations for inter-individual differences in the effectiveness of the internalization activities. We invited children via email to participate in this follow-up assessment, which took about 15 minutes to complete. Sixty-five children accepted this invitation. Children received a 5 Euro gift card for an online shop as compensation. Both performance tasks were programmed in PsychoPy (Peirce et al., 2019) and run on the Pavlovia server (https://pavlovia.org). The tasks, including the stimuli and stimulus lists, can be found at
https://osf.io/sbcfd/?view_only=cd294d0742a7422a99a460fd22541388.

Analogical Reasoning. We created a computerized version of the analogies subtest of the SON-R 6-40 (Tellegen et al., 2012). The test involved an adaptive testing procedure in which a maximum of 36 points could be achieved. One child was excluded from analyses involving analogical reasoning (pairwise exclusion) because it had a score of only 1 and had pressed the same key six times in a row.

Executive Functions. We assessed children's executive functioning (i.e., inhibition and switching) and mental processing speed with a slightly modified version of the Hearts and Flowers Task (HFT; (Wright & Diamond, 2014); for a description of the task parameters and testing procedure used here, see Brod et al., 2017).

We computed efficiency scores for inhibition and switching using the formula $1 - [(y - x) / x]$ (see Breitwieser & Brod, 2021). For the inhibition score, y denotes the RT divided by the accuracy in the incongruent block; for the switching score, y denotes the RT divided by the accuracy for congruent trials in the mixed block. For both scores, x denotes the RT divided by the accuracy in the congruent block. The latter was also used as a score for mental processing speed. Seven children's average RTs lay above the 1.5 Tukey fence and were thus excluded from all analyses involving scores from the HFT (pairwise exclusion).

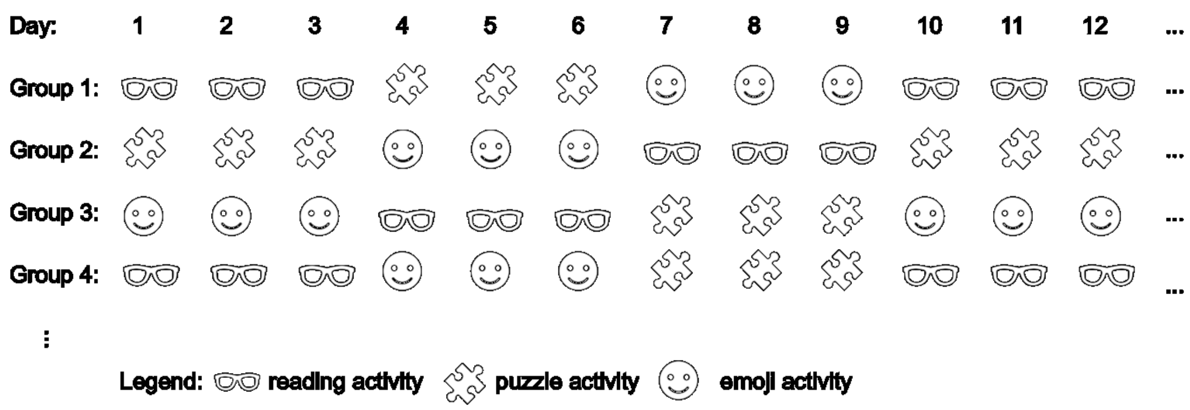
2.3 Design

Children were presented with one of three activities (emoji, puzzle, or reading) to internalize one plan per day. The same internalization activity was used for three consecutive days, after which a different activity was performed for three days, and so on (see Figure 2). Thus, children who completed the study learned 27 plans and performed each internalization activity nine times over the course of the study (within-subject design). We chose to present each internalization activity on three consecutive days so that the children had time to familiarize themselves with the activity and form an opinion about it before reporting their attitude toward using the activity (see section 2.2.5).

We counterbalanced the order of the internalization activities to control for sequence effects (i.e., one-third of the children started with the emoji activity, one-third with the puzzle activity, and one-third with the reading activity; see Figure 2). We also counterbalanced the order of the plans that the children had to internalize (i.e., the same plan was internalized by some children with the emoji activity, by others with the puzzle or reading activity). In total, there were six groups with different orders of internalization activities and plans. To ensure balanced group sizes, children were assigned to groups in the order of their registration (i.e., the first child was assigned to group 1, the second to group 2, and so on).

Figure 2

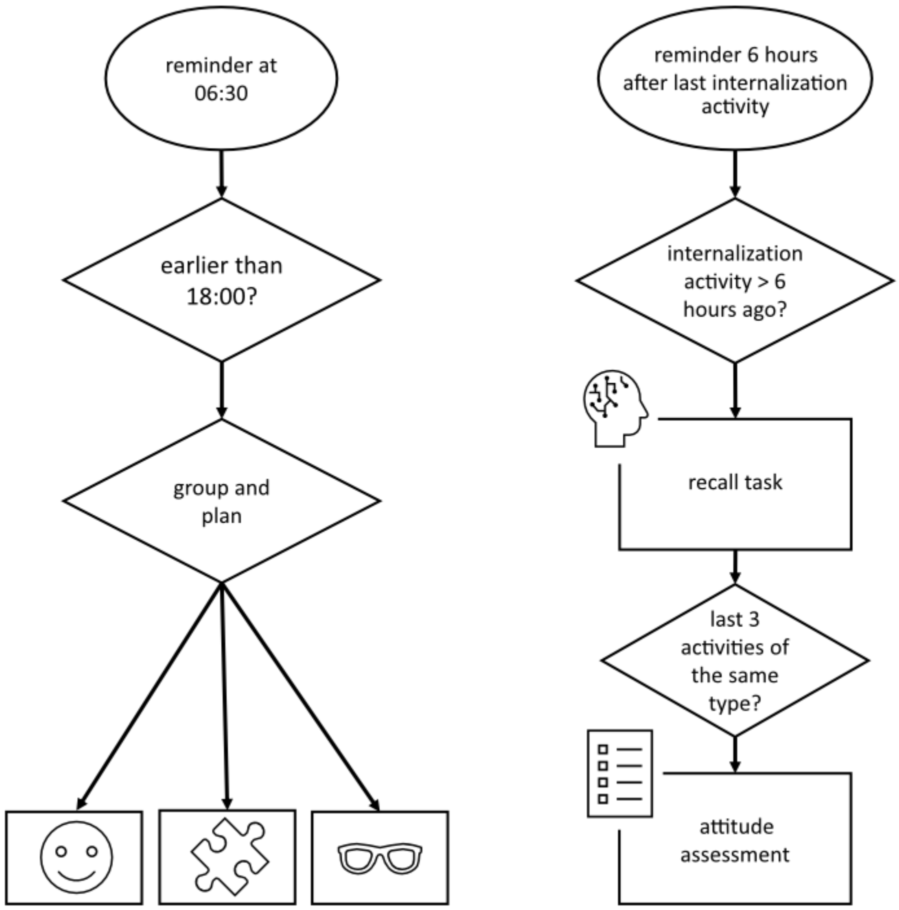
Schematic Overview of the Study Design



2.4 Procedure

Figure 3

Overview of the Procedure



After installing the app, the children were asked to complete several questionnaires, and received detailed instructions about the study and the three internalization activities. The study proper began on the day after the initial session (see Figure 3 for an overview of the procedure). From the first day on, the app sent a reminder notification every morning at 6:30 a.m. to remind the children of the internalization activity which had to be completed before 6 p.m. This deadline was chosen to have at least a six hour break between the internalization activity and the recall task without a night of sleep between both tasks, as we expected sleep to significantly affect memory performance (Scullin & McDaniel, 2010). Therefore, the reminder notification to perform the recall task appeared six hours after the completion of the internalization activity. If children attempted to use the app outside of these time constraints, the app displayed a message with instructions on the proper times to use the app for the study.

2.5 Data Analysis

The data and analysis scripts can be found on the Open Science Framework (https://osf.io/sbcfd/?view_only=cd294d0742a7422a99a460fd22541388). The alpha level was set at .05 throughout the analyses. For the exploratory analyses, we report *p*-values < .10 as "trends".

The pre-registered analyses were performed in R (R Core Team, 2021), using the package "lme4" (version 1.1-27.1, Bates et al., 2015) to compute mixed-effects models. We used a logistic mixed-effects model to test the effect of internalization activity (three levels: emoji, puzzle, reading; dummy-coded with "puzzle" as the reference category) on recall performance. The model further included random participant slopes and a random stimulus intercept.

We included all children in the analyses, regardless of whether they provided data from 27 or fewer recalls. Thus, the maximum likelihood estimation was based on the full information of the sample (without imputation). The significance of the fixed effect of internalization activity was tested with a likelihood-ratio test between the full model and a model without the fixed effect. Wald tests were performed via Satterthwaite's degrees of freedom method, using the package "lmerTest" (version 3.1-3, Kuznetsova et al., 2017).

Differences in children's attitude toward using the three internalization activities were explored using a linear mixed-effects model. The dependent variable was recall performance averaged across the three recalls prior to each attitude rating. The fixed and random effects structure was the same as in equation 1, except that there was no random stimulus intercept..

The moderating effects of the three time on task scores were explored by simultaneously entering the three interaction terms into the logistic mixed-effects model of equation 1. The random participant slopes were removed because the model did not converge when including them. The moderating effects of grade level and cognitive abilities were tested similarly.

The mediation analyses were carried out in Mplus (Muthén & Muthén, 2017). In the first path model, we regressed time on task in the puzzle condition on grade level (path a) and regressed recall performance in the puzzle condition on time on task in the puzzle condition (path b) and on grade level (path c′). In the second path model, we regressed time on task in the emoji condition on analogical reasoning (path a) and regressed recall performance in the emoji condition on time on task in the emoji condition (path b) and on analogical reasoning (path c′). The indirect effects of both models were calculated by multiplying paths a and b. The significance of the indirect effects were determined by calculating the confidence intervals via bias-corrected bootstrapping (MacKinnon et al., 2004).

3 Results

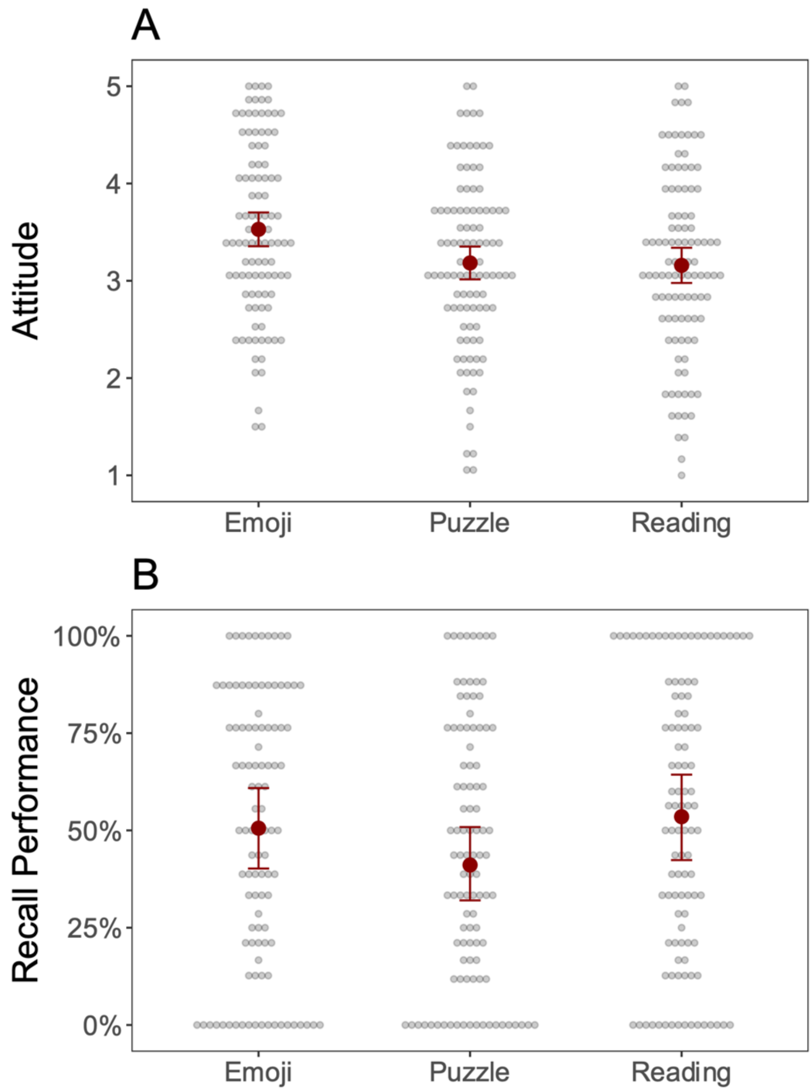
3.1 Children's Attitude Toward Using the Internalization Activities

The likelihood-ratio test revealed that children had differing attitudes toward the three activities ($\chi^2(2) = 17.61$, $p < .001$; see Table S1 in the supplement). Children's attitude toward using the emoji activity was significantly more positive ($M = 3.52$, $SD = 0.89$) than their attitude toward using the puzzle activity ($M = 3.17$, $SD = 0.88$; see Fig. 4a) or the reading activity ($M = 3.16$, $SD = 0.93$; $b = 0.37$, 95%-CI [0.20, 0.54], $p < .001$).

In the final questionnaire, 29.67% of the children reported that, if they had to again internalize plans every day, they would prefer to use the emoji activity. The second most popular option was to alternate between all three internalization activities, just as they did during the study (16.48%), followed by only using the puzzle activity (15.38%). Preferences for alternating between any two of the internalization activities ranged between 7.69% and 13.19%. The least popular option was to only re-read the plans (5.49%). A goodness-of-fit test confirmed that the probability to choose an option differed between the response options ($\chi^2(6) = 23.54$, $p = .001$). Taken together, these results suggest that children preferred using the emoji activity.

Figure 4

The Effects of Internalization Activity on Attitude (A) and Recall Performance (B)



Note. Gray dots = raw data points. Red dots = predicted values with 95% confidence interval.

3.2 The Effect of Internalization Activity on Children's Recall of the Plans

Children's average recall performance for the plans was about 50% (M = 48.96%, SD = 30.08%), indicating that children were able to successfully memorize a substantial proportion of plans but that the task was not an easy one. Our hypotheses regarding differences in recall performance between the three internalization activities (emoji > puzzle > reading) were only partly confirmed (see Table S1 in the supplement). The likelihood-ratio test revealed that children's recall performance differed between the internalization activities ($\chi^2(2) = 13.34$, $p = .001$). In accordance with hypothesis H1a, children were 1.47 times more likely to correctly recall the plans when they had internalized them with the emoji activity rather than the puzzle activity (Fig. 4b). However, contrary to hypothesis H1b, children were also 1.65 times more likely to correctly recall the plans when they had internalized them with the reading activity rather than the puzzle activity. There was no significant difference between the emoji activity and the reading activity ($b = -0.02$, 95%-CI [-0.07, 0.04], $p = 528$).

Taken together, when looking at average effects across children, the results were not in line with our hypothesis that internalizing the plans with an activity that promotes active or constructive engagement (i.e., puzzle, emoji) leads to better recall of the plans than an activity that promotes passive engagement (i.e., reading). Therefore, we next explored whether children actually engaged in the cognitive processes that we intended to stimulate with the active and constructive internalization activities.

Table 1

Descriptive Statistics and Correlations for Study Variables

	M (*SD*)	1	2	3	4	5	6	7	8	9	10
1) $\text{Recall}_{\text{Emoji}}$	49.42 (34.42)										
2) $\text{Recall}_{\text{Puzzle}}$	42.85 (33.02)	$.63^{***}$									
3) $\text{Recall}_{\text{Reading}}$	52.93 (35.74)	$.62^{***}$	$.66^{***}$								
4) $\text{ToT}_{\text{Emoji}}$	76.89 (34.72)	$.29^{**}$	$.21^{*}$	.16							
5) $\text{ToT}_{\text{Puzzle}}$	29.25 (7.98)	-.02	$.24^{*}$	.08	$.28^{**}$						
6) $\text{ToT}_{\text{Reading}}$	17.66 (1.71)	.10	-.01	.00	$.45^{***}$	$.22^{*}$					
7) Grade Level	/	.03	$-.17^{\dagger}$	-.08	-.02	$-.19^{\dagger}$	.01				

8) Analogical Reasoning	22.72 (6.41)	$.39^{**}$	$.28^{*}$	$.35^{**}$	$.33^{**}$	.04	$.35^{**}$	-.08			
9) Processing Speed	0.46 (0.08)	.02	.12	.12	$.24^{\dagger}$	.17	.11	$-.28^{*}$	.17		
10) Inhibition	0.72 (0.26)	$-.23^{\dagger}$	-.14	-.19	.11	.12	-.05	-.08	.03	$.43^{***}$	
11) Switching	-0.20 (0.54)	.00	-.01	-.06	$.29^{*}$	.22	.11	$-.24^{\dagger}$	.17	$.40^{**}$	$.44^{***}$

Note. Recall = Recall Performance, ToT = Time on Task.

$^{\dagger}p < .10$ $^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$

3.3 Time on Task

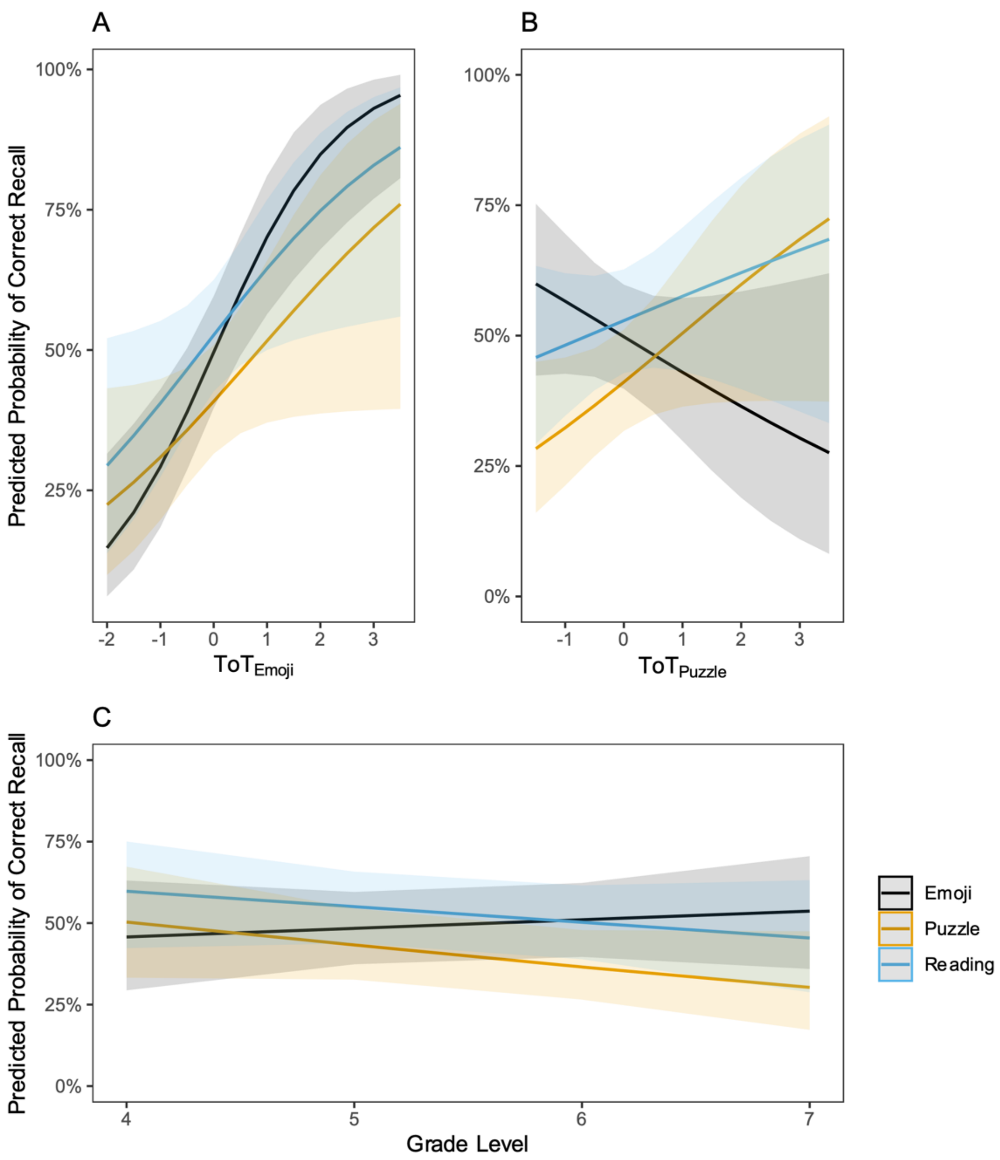
We used the time children spent performing the internalization activities as a proxy for the actual effort children invested when performing the internalization activities. For each child, we calculated the average time they spent performing each internalization activity. It should be noted that time on task is not equally meaningful for all activities, because the reading activity was timed, requiring children to spend at least 15 seconds on the internalization screen (see section 2.2.3). Nevertheless, we included time on task in the reading condition in the analyses.

Children's recall performance in the emoji condition was significantly correlated with their time spent performing the emoji activity (see Table 1). Recall performance in the puzzle condition was correlated with time spent performing the puzzle activity but also with time spent performing the emoji activity. Recall performance in the reading condition was not correlated with time spent performing the reading activity. These results suggest that children who spent more time performing the emoji and puzzle activity benefited more from these activities.

Next, we tested whether the average time children spent performing each activity predicted the *relative* effectiveness of the internalization activities (i.e., moderator effect of time on task; see Table S2 in the supplement). The likelihood-ratio tests revealed that two of the three interaction terms were significant: The relative effectiveness of the emoji activity increased for children who spent more time performing this activity ($\chi^2(2) = 8.19$, $p = .017$; see Fig. 5a). Similarly, the relative effectiveness of the puzzle activity increased for children who spent more time performing this activity ($\chi^2(2) = 17.45$, $p < .001$; see Fig. 5b). The relative effectiveness of the reading activity was not moderated by the time children spent performing this activity ($\chi^2(2) = 1.35$, $p = .508$). Taken together, these results suggest that the effectiveness of the active and constructive internalization activities were predicted by how much time children spent actually engaging with these activities.

Figure 5

Moderating Effects of Time on Task and Grade Level



Note. Scores of time on task are z-standardized.

3.4 Children's Prerequisites

So far, the findings suggest that different children benefited from different internalization activities depending on how much time they spent engaging with the activities (see section 3.3). Next, we explored whether differences in children's prerequisites predicted which internalization activity worked best for them. Specifically, we focused on children's grade level as a proxy for developmental factors as well as cognitive abilities.

3.4.1 Grade Level

The majority of children attended grades 5 (35.84%) and 6 (28.30%). About one sixth of the children attended grades 4 (16.98%) and 7 (16.04%), respectively. Grade level was not correlated with average recall performance in the emoji or reading condition, whereas there was a marginally significant negative correlation between grade level and recall performance in the puzzle condition (see Table 1). There was further a marginally significant negative correlation between grade level and time on task in the puzzle condition. Thus, children in higher grades spent less time performing the puzzle activity and had poorer recall of plans than children in lower grades when they used the puzzle activity.

Grade level was a significant moderator of the effect of internalization activity on recall performance ($\chi^2(2) = 7.97$, $p = .019$; see Table S3 in the supplement): Children in higher grades had poorer recall of the plans when using the puzzle or reading activity instead of the emoji activity (see Fig. 5c). Since grade level was also correlated with time on task in the puzzle condition, we calculated the indirect effect of grade level via time on task on recall performance in the puzzle condition (see Figure 6a). The indirect effect was significant, suggesting that children in higher grades spent less time performing the puzzle activity, which in turn predicted poorer recall of plans when using this activity.

3.4.2 Cognitive Abilities

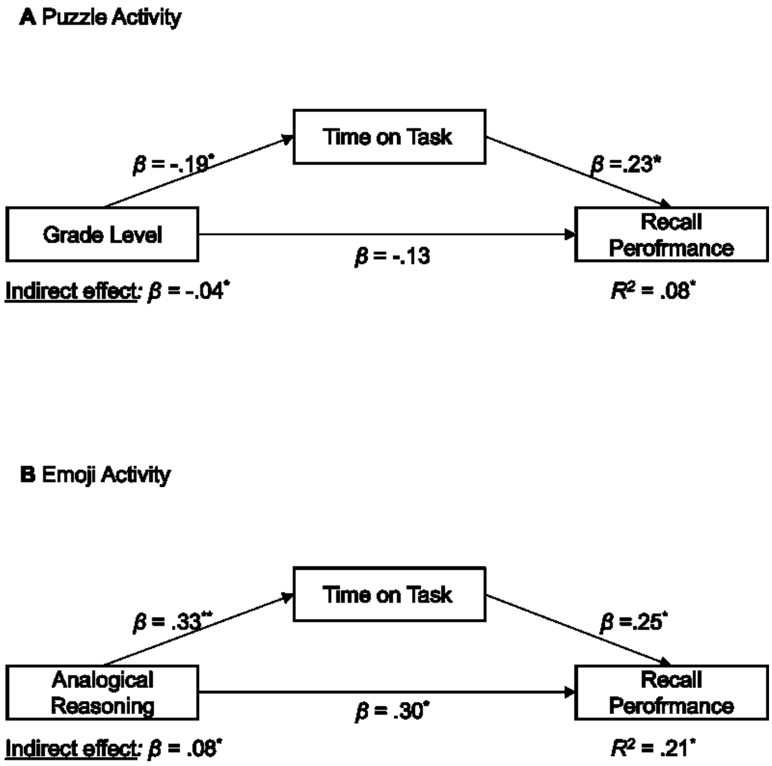
Processing speed, inhibition, and switching correlated positively among each other but not with analogical reasoning abilities (see Table 1). Analogical reasoning abilities correlated positively with recall performance in all conditions. Among the executive functions, only inhibition had a trend-level (but negative) correlation with recall performance in the emoji condition.

We entered all cognitive abilities as separate moderators of the effect of internalization activity on recall performance into the logistic mixed-effects model. None of the four interaction terms were significant (all $p > .10$; see Table S4 in the supplement). Thus, there was no evidence that children with higher cognitive abilities benefited more from a particular internalization activity than from another.

However, since analogical reasoning correlated with time on task in the emoji condition only (see Table 1), we tested whether, within the emoji condition, the association between analogical reasoning and recall performance was mediated by time on task (see Fig. 6b). The indirect effect was significant, suggesting that children with better analogical reasoning abilities spent more time when performing the emoji activity, which in turn predicted better recall of plans when using this activity.

Figure 6

Mediation Models



$^{*}p < .05$ $^{**}p < .01$

4 Discussion

Despite the many distractions in everyday life, remembering and implementing a plan is no easy task. Therefore, the present study tested how a planning app should best be designed to help children internalize plans. The different internalization activities were based on the ICAP framework that predicts which types of engagement should lead to deeper or shallower levels of cognitive processing (Chi et al., 2018; Chi & Wylie, 2014). In contrast to our hypotheses, internalizing a plan with the constructive emoji activity (i.e., creating a representation of the plan with emojis) was not generally superior to the passive activity of reading the plan three times. However, the newly developed emoji activity is a viable alternative to the more conventional re-reading activity, especially because children had a more positive attitude toward using it - an important determinant for technology adoption. Furthermore, our findings demonstrate that, in order to be child-appropriate, a planning app needs to be tailored to children's individual characteristics. Differences in grade level and analogical reasoning abilities predicted the time children spent performing the puzzle and emoji activity, respectively. Time on task, in turn, was an important predictor of the relative effectiveness of these activities. We discuss the implications of these findings for the development of a child-appropriate planning app.

4.1 Internalization Activities Need to Stimulate Deep Processing

Across the different activities, our findings highlight the need to support children in their planning. Even with the help of our app, children were able to recall the plans correctly only about 50% of the time. While we do not know what children's recall performance would have been without any assistance, research on metamemory abilities suggests that children are generally not particularly good at choosing effective learning strategies on their own (Bjorklund, 2007; Schneider, 2010). Thus, there should be large room for interventions to improve children's internalization of plans.

However, contrary to our hypotheses, our results suggest that prompting children to use deeper processing strategies does not necessarily improve children's recall performance. What could be reasons for this unexpected finding? Part of the answer could be that the low complexity of the task was not sensitive to the effects of deeper processing activities. The authors of the ICAP framework proposed that deeper processing strategies have less influence in simple retrieval tasks than in more complex knowledge change tasks (Chi & Wylie, 2014). However, the effects of retrieval practice, which we aimed to elicit with the puzzle activity, have been primarily studied and evidenced with simple learning tasks such as the retrieval of vocabulary words (Fritz et al., 2007; Goossens et al., 2014). Furthermore, previous research found positive effects of vivid imagination - a deeper processing strategy - compared to passive reading for implementation intentions (Knäuper et al., 2009). We therefore assume that task characteristics were not the primary reason that, across children, deeper internalization activities did not improve the recall of plans compared to passive re-reading.

Rather, our findings suggest that a crucial factor was the time children spent performing the deeper internalization activities (i.e., puzzle, emoji). The effectiveness of the constructive internalization activity increased with the time children spent performing the activity. Although not a perfect measure, time on task reflects processing depth (Craik & Lockhart, 1972). That is, elaborating on the meaning of the plan and searching for emojis that represent this meaning takes more time than quickly or even randomly selecting emojis. It appears that some children were not willing or able to engage in deeper processing, at least not to the extent needed to improve performance compared to re-reading.

A similar conclusion can be drawn for the puzzle activity: this activity was only effective for children who spent more time performing it. However, the puzzle activity differed from the emoji activity in that the outcome of this activity (i.e., the unscrambled sentence) had a right or wrong answer, and children could only continue when they had arrived at the correct answer. Thus, children who arrived at the correct answers more quickly benefitted less from this activity than children who needed more time. Taken together, our results suggest that the relative effectiveness of different internalization activities does not primarily depend on the characteristics of the activity. Instead, the relative effectiveness of deeper internalization activities depends primarily on how children perform the activity. Therefore, the difficulty of stimulating deep internalization of plans lies in designing activities that children are both willing and able to perform effectively.

4.2 Taking Children's Prerequisites Into Account When Developing a Child-Appropriate Planning App

Our findings suggest that no one internalization activity is most effective for every child. Rather, children differ in what activity is most effective in helping them internalize plans. Therefore, the main take-away from our study is that, to be appropriate for every child, a planning app should be personalized. Personalized technology caters to the needs and characteristics of the individual learner (Xie et al., 2019). It is thus crucial to identify individual prerequisites that predict the differential effectiveness of different technology-enhanced internalization activities.

Our study provides initial evidence that developmental factors play a role in which internalization activities are effective for which child. Compared to children in lower grades, children in higher grades performed worse when using the re-reading or puzzle activity. The fact that children in higher grades spent less time performing the puzzle activity, which in turn predicted recall performance, suggests that the activity was too easy for these children. One explanation is that older children might have been able to unscramble the sentences without retrieving the plan from memory, simply by putting the individual words together to form a coherent sentence. Alternatively, older children might have retrieved the sentence from memory, but so effortlessly that it had no memory-enhancing effect (Bjork & Bjork, 2011). Although the specific cognitive or psychological mechanisms underlying the effect cannot be determined at this point, it shows that selecting age-appropriate internalization activities for planning is important. Learning strategy research has uncovered many such age-related individual differences in the effectiveness of learning strategies (e.g., Breitwieser & Brod, 2021). However, up until now, an individual differences approach has not yet been applied to research on internalization strategies for planning.

Research suggests that differences in cognitive abilities moderate the effectiveness of learning strategies in children (Brod, 2021). The cognitive abilities explored in the present study did not moderate the effectiveness of the internalization activities. Executive functions and processing speed were unrelated to recall performance, while analogical reasoning positively predicted recall performance regardless of the internalization activity. Nevertheless, it is still worthwhile to consider cognitive abilities when selecting internalization activities for children. Analogical reasoning was specifically related to how much time children spent performing the emoji activity. It could be argued that the emoji activity requires children to engage in analogical reasoning by comparing different emojis among each other and to the meaning of the plans to decide which emojis fit best (cf. Duit, 1991). The mediating role of time on task suggests that children with lower analogical reasoning abilities might be less inclined to perform these effortful cognitive operations. However, *if* they perform them, these children can also benefit from the emoji activity. These considerations demonstrate the added value of investigating the cognitive mechanisms of various internalization activities to identify children that may need more support in performing them.

4.3 Limitations & Directions for Future Research

The present study has several implications that should be kept in mind when interpreting the results. First, we did not test whether better recall of a plan is actually related to the effectiveness of the plan. Although there is still debate on the mechanisms underlying prospective memory, some researchers suggest that prospective memory relies on automatic retrieval processes (see Kvavilashvili et al., 2008). In the specific case of implementation intentions, it is assumed that execution of a plan is triggered without conscious intent (Bayer et al., 2009; Gollwitzer, 1999). However, this automation process may still rely on conscious retrieval processes (McDaniel & Einstein, 2000). Future research should test whether better conscious recall of a plan predicts better execution of the plan.

Moving forward, we recognize the need for a more in-depth investigation of the emoji activity to determine how well the inputs truly correspond to the plans. As this activity encouraged creativity and personal associations, it can be challenging for third parties to understand if an emoji representation accurately reflects a plan. Future studies could explore whether there is a relationship between the selected emojis and the ability to remember the plans. Ideally, such investigations should take place in a laboratory setting where researchers can observe and understand the thought processes of the children. This could lead to a better understanding of the underlying mechanisms of the emoji activity and help to optimize its effectiveness in future interventions. Another potential issue concerns the plans that we presented to the children. We did not create plans that we considered particularly effective because our focus was on recall performance. As the children internalized many of the plans without an intention of actually using them, the motivation was likely lower than for highly relevant plans. This, in turn, may have had a negative impact on children's motivation to use the internalization activities effectively. Therefore, we see a clear need to replicate the current study in a context where children create and internalize their own plans.

There are also limitations with regard to the follow-up assessment of cognitive abilities. First, only a subsample of children participated in the follow-up assessment. Therefore, this subsample might be selective (e.g., children who are conscientious and enjoyed the previous study). Second, a broader assessment of cognitive abilities would be necessary to determine which abilities are specifically relevant for which internalization activity.

Lastly, many of our findings are based on exploratory and correlational analyses. Therefore, we cannot draw causal conclusions from them. For instance, our findings suggest that time on task potentially mediates the effects of analogical reasoning abilities on recall performance in the emoji condition. To properly test this assumption, future research could manipulate time on task to see if it has a causal effect on the effectiveness of the emoji activity.

4.4 Conclusion

The goal of this study was to develop a child-appropriate planning app that helps children make and internalize plans in their daily lives. We specifically focused on an important yet often neglected part of planning: How to internalize plans deeply so that they can be retrieved in everyday life. To ensure high ecological validity, we tested the effectiveness of the planning app in a naturalistic setting that mimics this challenge. That is, children used the app in their daily lives over an extended period of time, while children's usage of the app-based interventions could be measured objectively via logfiles. In sum, our findings reveal that no one app-based internalization activity will be the most effective intervention for every child. Therefore, it is paramount that the development of a *personalized* planning app is based on an understanding of the individual characteristics that influence the effectiveness of planning intervention. This study lays the groundwork for this endeavor.

Disclosure Statement

The authors report there are no competing interests to declare.

References

Achtziger, A., Gollwitzer, P. M., & Sheeran, P. (2008). Implementation intentions and shielding goal striving from unwanted thoughts and feelings. *Personality and Social Psychology Bulletin*, *34*(3), 381–393. https://doi.org/10.1177/0146167207311201

Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Prentice Hall.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036–1060. https://doi.org/10.1037/0033-295X.111.4.1036

Aslan, A., & Bäuml, K.-H. T. (2016). Testing enhances subsequent learning in older but not in younger elementary school children. *Developmental Science*, *19*(6), 992–998. https://doi.org/10.1111/desc.12340

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bayer, U. C., Achtziger, A., Gollwitzer, P. M., & Moskowitz, G. B. (2009). Responding to subliminal cues: Do if-then plans facilitate action preparation and initiation without conscious intent? *Social Cognition*, *27*(2), 183–201. https://doi.org/10.1521/soco.2009.27.2.183

Bjork, R. A., & Bjork, E. L. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the Real World: Essays Illustrating Fundamental Contributions to Society* (pp. 56–64). Worth Publishers.

Bjorklund, D. F., Dukes, C., & Brown, R. D. (2008). The development of memory strategies. In M. Courage & N. Cowan (Eds.), *The Development of Memory in Infancy and Childhood* (pp. 157–188). Psychology Press.

https://doi.org/https://doi.org/10.4324/9780203934654

Brandstätter, V., Lengfelder, A., & Gollwitzer, P. M. (2001). Implementation intentions and efficient action initiation. *Journal of Personality and Social Psychology*, *81*(5), 946–960. https://doi.org/10.1037/0022-3514.81.5.946

Breitwieser, J., & Brod, G. (2021). Cognitive prerequisites for generative learning: Why some learning strategies are more effective than others. *Child Development, 92*(1), 258–272. https://doi.org/10.1111/cdev.13393

Breitwieser, J., Neubauer, A. B., Schmiedek, F., & Brod, G. (2022). Self-regulation prompts promote the achievement of learning goals – But only briefly: Uncovering hidden dynamics in the effects of a psychological intervention. *Learning and Instruction, 80*, 101560. https://doi.org/10.1016/j.learninstruc.2021.101560

Brewer, R. N., Morris, M. R., & Lindley, S. E. (2017). How to remember what to remember: Exploring possibilities for digital reminder systems. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *1*(3), 1–20. https://doi.org/10.1145/3130903

Brod, G. (2021). Generative learning: Which strategies for what age? *Educational Psychology Review, 33*(4), 1295–1318. https://doi.org/10.1007/s10648-020-09571-9

Brod, G., Bunge, S. A., & Shing, Y. L. (2017). Does one year of schooling improve children's cognitive control and alter associated brain activation? *Psychological Science, 28*(7), 967–978. https://doi.org/10.1177/0956797617699838

Chi, M. T. H., Adams, J., Bogusch, E. B., Bruchok, C., Kang, S., Lancaster, M., Levy, R., Li, N., McEldoon, K. L., Stump, G. S., Wylie, R., Xu, D., & Yaghmourian, D. L. (2018). Translating the ICAP theory of cognitive engagement into practice. *Cognitive Science*, *42*(6), 1777–1832. https://doi.org/10.1111/cogs.12626

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Cho, K.-S., & Lee, J.-M. (2017). Influence of smartphone addiction proneness of young children on problematic behaviors and emotional intelligence: Mediating self-assessment effects of parents using smartphones. *Computers in Human Behavior*, *66*, 303–311. https://doi.org/10.1016/j.chb.2016.09.063

Cohen, A.-L., Bayer, U. C., Jaudas, A., & Gollwitzer, P. M. (2007). Self-regulatory strategy and executive control: implementation intentions modulate task switching and Simon task performance. *Psychological Research*, *72*(1), 12–26. https://doi.org/10.1007/s00426-006-0074-2

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science*, *35*(8), 982–1003. https://doi.org/10.1287/mnsc.35.8.982

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Doherty, K., & Doherty, G. (2019). Engagement in HCI. *ACM Computing Surveys*, *51*(5), 1–39. https://doi.org/10.1145/3234149

Drachsler, H., Jansen, J., & Kirschner, P. A. (2021). Adoption of learning technologies in times of pandemic crisis. *Journal of Computer Assisted Learning, 37*(6), 1509–1512. https://doi.org/10.1111/jcal.12626

Duit, R. (1991). On the role of analogies and metaphors in learning science. *Science Education*, *75*(6), 649–672. https://doi.org/10.1002/sce.3730750606

Fennis, B. M., Adriaanse, M. A., Stroebe, W., & Pol, B. (2011). Bridging the intention-behavior gap: Inducing implementation intentions through persuasive appeals. *Journal of Consumer Psychology*, *21*(3), 302–311. https://doi.org/10.1016/j.jcps.2010.12.003

Fiorella, L., & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational Psychology Review*, *28*(4), 717–741. https://doi.org/10.1007/s10648-015-9348-9

Fritz, C. O., Morris, P. E., Nolan, D., & Singleton, J. (2007). Expanding retrieval practice: An effective aid to preschool children's learning. *Quarterly Journal of Experimental Psychology*, *60*(7), 991–1004. https://doi.org/10.1080/17470210600823595

Gollwitzer, P. M. (1990). Action phases and mind-sets. In E. T. Higgins & R. M. Sorrentino (Eds.), *The handbook of motivation and cognition: Foundations of social behavior* (Vol. 2, pp. 53–92). Guilford Press.

Gollwitzer, P. M. (1996). The volitional benefits of planning. In P. M. Gollwitzer & J. A. Bargh (Eds.), *The psychology of action: Linking cognition and motivation to behavior* (pp. 287–312). Guilford Press.

Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *American Psychologist*, *54*(7), 493–503. https://doi.org/10.1177/0146167207311201

Goossens, N. A. M. C., Camp, G., Verkoeijen, P. P. J. L., & Tabbers, H. K. (2014). The effect of retrieval practice in primary school vocabulary learning. *Applied Cognitive Psychology*, *28*(1), 135–142. https://doi.org/10.1002/acp.2956

Hagger, M. S., Luszczynska, A., de Wit, J., Benyamini, Y., Burkert, S., Chamberland, P. E., Chater, A., Dombrowski, S. U., van Dongen, A., French, D. P., Gauchet, A., Hankonen, N., Karekla, M., Kinney, A. Y., Kwasnicka, D., Hing Lo, S., López-Roig, S., Meslot, C., Marques, M. M., … Gollwitzer, P. M. (2016). Implementation intention and planning interventions in Health Psychology: Recommendations from the Synergy Expert Group for research and practice. *Psychology and Health*, *31*(7), 814–839.

https://doi.org/10.1080/08870446.2016.1146719

Hoch, E., Scheiter, K., & Schüler, A. (2020). Implementation intentions for improving self-regulation in multimedia learning: Why don't they work? *The Journal of Experimental Education*, *88*(4), 536–558. https://doi.org/10.1080/00220973.2019.1628693

Jacob, L., Lachner, A., & Scheiter, K. (2022). Do school students' academic self-concept and prior knowledge constrain the effectiveness of generating technology-mediated explanations? *Computers & Education*, *182*, 104469. https://doi.org/10.1016/j.compedu.2022.104469

Johannes, N., Veling, H., Verwijmeren, T., & Buijzen, M. (2019). Hard to resist? The effect of smartphone visibility and notifications on response inhibition. *Journal of Media Psychology*, *31*(4), 214–225. https://doi.org/10.1027/1864-1105/a000248

Knäuper, B., Roseman, M., Johnson, P. J., & Krantz, L. H. (2009). Using mental imagery to enhance the effectiveness of implementation intentions. *Current Psychology*, *28*(3), 181–186. https://doi.org/10.1007/s12144-009-9055-0

Kramer, J.-N., Künzler, F., Mishra, V., Smith, S. N., Kotz, D., Scholz, U., Fleisch, E., & Kowatsch, T. (2020). Which components of a smartphone walking app help users to reach personalized step goals? Results from an optimization trial. *Annals of Behavioral Medicine*, *54*(7), 518–528. https://doi.org/10.1093/abm/kaaa002

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Kvavilashvili, L., Kyle, F., & Messer, D. J. (2008). The development of prospective memory in children: Methodological issues, empirical findings, and future directions. In M. Kliegel, M. A. McDaniel, & G. O. Einstein (Eds.), *Prospective memory: Cognitive, neuroscience, developmental, and applied perspectives* (pp. 115–140). Taylor & Francis

Group/Lawrence Erlbaum Associates. https://doi.org/10.1016/j.dr.2014.08.001

Lewis, C. (1982). Using the "thinking aloud" method in cognitive interface design. In *Ibm. Rc-9265*.

Lewis, J. R. (1994). Sample sizes for usability studies: Additional considerations. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *36*(2), 368–378. https://doi.org/10.1177/001872089403600215

Lozano-Blasco, R., Robres, A. Q., & Sánchez, A. S. (2022). Internet addiction in young adults: A meta-analysis and systematic review. *Computers in Human Behavior*, *130*, 107201. https://doi.org/10.1016/j.chb.2022.107201

MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, *39*(1), 99–128. https://doi.org/10.1207/s15327906mbr3901_4

Mahapatra, S. (2019). Smartphone addiction and associated consequences: role of loneliness and self-regulation. *Behaviour & Information Technology*, *38*(8), 833–844. https://doi.org/10.1080/0144929X.2018.1560499

Martens, M. (2012). Issues of access and usability in designing digital resources for children. *Library and Information Science Research*, *34*(3), 159–168. https://doi.org/10.1016/j.lisr.2011.12.003

Martiny-Huenger, T., Bieleke, M., Oettingen, G., & Gollwitzer, P. M. (2016). From thought to automatic action: Strategic and spontaneous action control by if-then planning. In R. Deutsch, B. Gawronski, & W. Hofmann (Eds.), *Reflective and impulsive determinants of behavior* (pp. 69–84). Psychology Press.

McDaniel, M. A., & Einstein, G. O. (2000). Strategic and automatic processes in prospective memory retrieval: a multiprocess framework. *Applied Cognitive Psychology*, *14*(7), S127–S144. https://doi.org/10.1002/acp.775

Muthén, L. K., & Muthén, B. O. (2017). *Mplus User's Guide. Eighth Edition.* Muthén & Muthén.

Oettingen, G., Pak, H., & Schnetter, K. (2001). Self-regulation of goal setting: turning free fantasies about the future into binding goals. *Journal of Personality and Social Psychology*, *80*(5), 736–753. https://doi.org/10.1037/0022-3514.80.5.736

Oh, H. J., & Larose, R. (2015). Tell me a story about healthy snacking and i will follow: Comparing the effectiveness of self-generated versus message-aided implementation intentions on promoting healthy snacking habits among college students. *Health Communication*, *30*(10), 962–974. https://doi.org/10.1080/10410236.2014.910289

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pinder, C., Vermeulen, J., Wicaksono, A., Beale, R., & Hendley, R. J. (2016). If this, then habit. *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, 690–697. https://doi.org/10.1145/2957265.2961837

Pirolli, P., Mohan, S., Venkatakrishnan, A., Nelson, L., Silva, M., & Springer, A. (2017). Implementation intention and reminder effects on behavior change in a mobile health system: A predictive cognitive model. *Journal of Medical Internet Research*, *19*(11), e397. https://doi.org/10.2196/jmir.8217

R Core Team. (2021). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.r-project.org/

Schneider, W. (2010). The development of metacognitive competencies. In B. Glatzeder, V. Goel, & A. Müller (Eds.), *Towards a Theory of Thinking. On Thinking.* (pp. 203–214). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-03129-8_14

Scullin, M. K., & McDaniel, M. A. (2010). Remembering to execute a goal: Sleep on it! *Psychological Science*, *21*(7), 1028–1035. https://doi.org/10.1177/0956797610373373

Smith, A. L., & Chaparro, B. S. (2015). Smartphone text input method performance, usability, and preference with younger and older adults. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *57*(6), 1015–1028. https://doi.org/10.1177/0018720815575644

Stawarz, K., Cox, A. L., & Blandford, A. (2015). Beyond self-tracking and reminders: Designing smartphone apps that support habit formation. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 2653–2662. https://doi.org/10.1145/2702123.2702230

Tabuenca, B., Kalz, M., Drachsler, H., & Specht, M. (2015). Time will tell: The role of mobile learning analytics in self-regulated learning. *Computers & Education, 89*, 53–74. https://doi.org/10.1016/j.compedu.2015.08.004

Tellegen, P. J., Laros, J. A., & Petermann, F. (2012). *Snijders-Oomen Nonverbal Intelligence Test (SON-R 6-40)*. Hogrefe.

Tukey, J. W. (1977). *Exploratory data analysis*. Addion-Wesely.

Veenman, M. V. J., Van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning*, *1*(1), 3–14. https://doi.org/10.1007/s11409-006-6893-0

Venkatesh, Morris, Davis, & Davis. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, *27*(3), 425. https://doi.org/10.2307/30036540

Virzi, R. A. (1992). Refining the test phase of usability evaluation: How many subjects is enough? *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *34*(4), 457–468. https://doi.org/10.1177/001872089203400407

Visuri, A., van Berkel, N., Okoshi, T., Goncalves, J., & Kostakos, V. (2019). Understanding smartphone notifications' user interactions and content importance. *International Journal of Human-Computer Studies*, *128*(February), 72–85. https://doi.org/10.1016/j.ijhcs.2019.03.001

Wang, G., Wang, Y., & Gai, X. (2021). A meta-analysis of the effects of mental contrasting with implementation intentions on goal attainment. *Frontiers in Psychology*, *12*(May), 1–10. https://doi.org/10.3389/fpsyg.2021.565202

Ward, A. F., Duke, K., Gneezy, A., & Bos, M. W. (2017). Brain drain: The mere presence of one's own smartphone reduces available cognitive capacity. *Journal of the Association for Consumer Research*, *2*(2), 140–154. https://doi.org/10.1086/691462

Webb, T. L., & Sheeran, P. (2007). How do implementation intentions promote goal attainment? A test of component processes. *Journal of Experimental Social Psychology*, *43*(2), 295–302. https://doi.org/10.1016/j.jesp.2006.02.001

Wicaksono, A., Beale, R., & Hendley, R. J. (2019). Using reinforced implementation intentions to support habit formation. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3290607.3312985

Wicaksono, A., Hendley, R., & Beale, R. (2019). Investigating the impact of adding plan reminders on implementation intentions to support behaviour change. *Interacting with Computers*, *31*(2), 177–191. https://doi.org/10.1093/iwc/iwz012

Wright, A., & Diamond, A. (2014). An effect of inhibitory load in children while keeping working memory load constant. *Frontiers in Psychology*, *5:213*, 1–9. https://doi.org/10.3389/fpsyg.2014.00213

Xie, H., Chu, H. C., Hwang, G. J., & Wang, C. C. (2019). Trends and development in technology-enhanced adaptive/personalized learning: A systematic review of journal publications from 2007 to 2017. *Computers and Education*, *140*, 103599. https://doi.org/10.1016/j.compedu.2019.103599

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, *41*(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2

B: Author's Contributions

Paper 1: Biedermann, D., Ciordas-Hertel, G., Winter, M., Mordel, J., and Drachsler, H. (2023). *Contextualized Logging of On-Task and Off-Task Behavior during Learning*. Accepted and in the copyediting phase in the Journal of Learning Analytics

Study conception and design: D.B., G.C.-H., M.W., and J.M.
Development of the data tracking system: D.B., G.C-H.
Research questions and formal analyses: D.B.
Draft manuscript preparation: D.B.
Critical commentary: G.C.-H., M.W., J.M. and H.D.
Funding and supervision: H.D.
All authors approved the final version of the paper.

Paper 2: Biedermann, D., Schneider, J., Ciordas-Hertel, G.-P., Eichmann, B., Hahnel, C., Goldhammer, F., & Drachsler, H. (2023). *Detecting the Disengaged Reader—Using Scrolling Data to Predict Disengagement during Reading*. LAK23: 13th International Learning Analytics and Knowledge Conference, 585–591. https://doi.org/10.1145/3576050.3576078

The study used data that was collected during the DiFA project, a cooperation between the DIPF | Leibniz-Institute for Research and Information and Education, the Goethe-University in Frankfurt, and the Leibniz-Institute for Educational Trajectories. All members of the DiFA team contributed to varying degrees to the design of the full one-semester moodle course.

Study conception and design: D.B.
Development of the data tracking system: D.B. and G.C-H.
Research questions and formal analyses: D.B.
Draft manuscript preparation: D.B.
Critical commentary: J.S., G.C.-H., B.E., C.H., F.G., and H.D.
Funding and supervision: H.D. and F.G.
All authors approved the final version of the paper.

Paper 3: Biedermann, D., Schneider, J., and Drachsler, H. (2021). *Digital self-control interventions for distracting media multitasking — A systematic review*. Journal of Computer Assisted Learning, 37(5), 1217–1231. https://doi.org/10.1111/jcal.12581

Study conception and design: D.B.
Data collection: D.B.
Coding: D.B. and J.S.
Research questions and formal analyses: D.B.
Draft manuscript preparation: D.B.
Critical commentary: J.S. and H.D.
Funding and supervision: H.D.
All authors approved the final version of the paper.

Paper 4: Biedermann, D., Kister, S., Breitwieser, J., Weidlich, J., and Drachsler, H. (2023). Use of Digital Self-Control Tools in Higher Education – A Survey Study. [Revisions submitted to the journal *Education and Information Technology*]

Study conception and design: D.B. and S.K.
Data collection: D.B.
Coding of qualitative data: D.B., S.K. and J.B.
Research questions and formal analyses: D.B. with feedback from J.B.
Draft manuscript preparation: D.B.
Critical commentary: S.K., J.B., J.W., and H.D.
Funding and supervision: H.D.
All authors approved the current revision of the paper.

Paper 5: Biedermann, D.*, Breitwieser, J*., Nobbe, L., Drachsler, H., and Brod, G. (2023). Designing an app to enhance children's planning skills: A case for personalized technology. [Revisions submitted to the journal Behavior and Information Technology]

The study was part of the PROMPT project, which is a project that aims to investigate digital prompting techniques in mobile apps for children.
Study conception and design: D.B., J.B., L.N., and G.B.
App development: D.B.

Data collection: D.B., J.B., and L.N.
Research questions and formal analyses: J.B. and D.B.
Draft manuscript preparation: J.B. and D.B.
Critical commentary: D.B., J.B., L.N., H.D. and G.B.
Funding and supervision: G.B. and H.D.
All authors approved the current revision of the paper.

Frankfurt am Main, den 03.08.2023

Daniel Biedermann

C: General Statements

Erklärung

Ich erkläre hiermit, dass ich mich bisher keiner Doktorprüfung unterzogen habe.

Frankfurt am Main, den 03.08.2023

(Unterschrift)

Eidesstattliche Versicherung

Ich erkläre hiermit, dass ich die vorgelegte Dissertation mit dem Titel:

"Technological Responses to Distracting Media Multitasking in Academic Learning"

selbstständig angefertigt und mich anderer Hilfsmittel als der in ihr angegebenen nicht bedient habe, insbesondere, dass alle Entlehnungen aus anderen Schriften mit Angabe der betreffenden Schrift gekennzeichnet sind.

Ich versichere, die Grundsätze der guten wissenschaftlichen Praxis beachtet, und nicht die Hilfe einer kommerziellen Promotionsvermittlung in Anspruch genommen zu haben.

Frankfurt am Main, den 03.08.2023

(Unterschrift)

D: List of Publications

Biedermann, D., Breitwieser, J., Nobbe, L., Drachsler, H., & Brod, G. (2023). *Designing an app to enhance children's planning skills: A case for personalized technology* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/ak3d7

Biedermann, D., Ochs, M., & Mester, R. (2015). COnGRATS: Realistic simulation of traffic sequences for autonomous driving. *2015 International Conference on Image and Vision Computing New Zealand (IVCNZ)*, 1–6. https://doi.org/10.1109/IVCNZ.2015.7761558

Biedermann, D., Ochs, M., & Mester, R. (2016). Evaluating visual ADAS components on the COnGRATS dataset. *2016 IEEE Intelligent Vehicles Symposium (IV)*, 986–991. https://doi.org/10.1109/IVS.2016.7535508

Biedermann, D., Schneider, J., Ciordas-Hertel, G.-P., Eichmann, B., Hahnel, C., Goldhammer, F., & Drachsler, H. (2023). Detecting the Disengaged Reader—Using Scrolling Data to Predict Disengagement during Reading. *LAK23: 13th International Learning Analytics and Knowledge Conference*, 585–591. https://doi.org/10.1145/3576050.3576078

Biedermann, D., Schneider, J., & Drachsler, H. (2021). Digital self-control interventions for distracting media multitasking - A systematic review. *Journal of Computer Assisted Learning*, *37*(5), 1217–1231. https://doi.org/10.1111/jcal.12581

Biedermann, D., Schwarz, P. O., Yau, J., & Drachsler, H. (2023). The Effect of Social Support Features via Buddies in App-Based Habit Building. *International Journal of Mobile and Blended Learning*, *15*(2), 1–12. https://doi.org/10.4018/IJMBL.318223

Amaefule, C. O., Breitwieser, J., **Biedermann, D.**, Nobbe, L., Drachsler, H., & Brod, G. (n.d.). Fostering children's acceptance of educational apps: The importance of designing enjoyable learning activities. *British Journal of Educational Technology*, *n/a*(n/a). https://doi.org/10.1111/bjet.13314

Ahmad, A., Schneider, J., Griffiths, D., **Biedermann, D.**, Schiffner, D., Greller, W., & Drachsler, H. (2022). Connecting the dots – A literature review on learning analytics indicators from a learning design perspective. *Journal of Computer Assisted Learning*, jcal.12716. https://doi.org/10.1111/jcal.12716

Biedermann, S. V., **Biedermann, D.**, Wenzlaff, F., Kurjak, T., Nouri, S., Auer, M. K., Wiedemann, K., Briken, P., Haaker, J., Lonsdorf, T. B., & Fuss, J. (2017). An elevated plus-maze in mixed reality for studying human anxiety-related behavior. *BMC Biology*, *15*(1), 125. https://doi.org/10.1186/s12915-017-0463-6

Biedermann, S. V., Roth, L., **Biedermann, D.**, & Fuss, J. (2022). Reliability of repeated exposure to the human elevated plus-maze in virtual reality: Behavioral, emotional, and autonomic responses. *Behavior Research Methods*. https://doi.org/10.3758/s13428-022-02046-5

Breitwieser, J., Nobbe, L., **Biedermann, D.**, & Brod, G. (2023). *Boosting self-regulated learning with mobile interventions: Planning and prompting help children maintain a regular study routine* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/e56tm

Ciordas-Hertel, G.-P., **Biedermann, D.**, Winter, M., Mordel, J., & Drachsler, H. (2022). How can Interaction Data be Contextualized with Mobile Sensing to Enhance Learning Engagement Assessment in Distance Learning? *INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION*, 105–112. https://doi.org/10.1145/3536220.3558037

Rödling, S., **Biedermann, D.**, Schneider, J., & Drachsler, H. (2020). Associative Media Learning With Smartwatches. *2020 IEEE Global Engineering Education Conference (EDUCON)*, 752–755. https://doi.org/10.1109/EDUCON45650.2020.9125157

E: Curriculum Vitae

Persönliche Informationen

Daniel Biedermann, M.Sc.
*15.01.1987 in Frankfurt am Main

Adresse: Markgrafenstraße 7, 60487 Frankfurt am Main
Telefon: +49 173 878 210 6
Email: dgbiedermann@gmail.com

Beruflicher Werdegang

Seit 11/2017	**DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation** Wissenschaftlicher Mitarbeiter Abteilung: Informationszentrum Bildung Betreuer: Prof. Dr. Hendrik Drachsler
02/2010 – 11/2017	**Dessoi GmbH** Softwareentwicklung und Consulting

Ausbildung

08/2015	**Master of Science Informatik** (⌀ 1,3) Goethe-Universität Frankfurt am Main Thema: „Simulation of Traffic Scenes using Computer Graphics Methods and Motion Models“ Betreuer: Prof. Dr. Rudolf Mester
03/2012	**Bachelor of Science Informatik** (⌀ 1,8) Goethe-Universität Frankfurt am Main Thema “Produktion Interaktiver Fragetypen mit Hilfe von Web-Standards“ Betreuer: Prof. Dr. Detlef Krömker
06/2006	**Abitur** Friedrich-Dessauer Gymnasium Frankfurt am Main