Development of Interactive Performance Measures for two Components of ICT Literacy: Successfully Accessing and Evaluating Information

Dissertation

zur Erlangung des

Doktorgrades der Naturwissenschaften

vorgelegt dem Fachbereich 05

Psychologie und Sportwissenschaften

der Johann Wolfgang Goethe-Universität

in Frankfurt am Main

von

Yvonne Nicole Keßel geb. Pfaff

aus Frankfurt am Main

Frankfurt am Main 2017

(D 30)

vom Fachbereich 05 der
Johann Wolfgang Goethe-Universität als Dissertation angenommen.

Dekan: Prof. Dr. Dr. Winfried Banzer

Gutachter: Prof. Dr. Frank Goldhammer
Prof. Dr. Johannes Hartig

Datum der Disputation:

Acknowledgements

I would like to take this opportunity to thank everyone who has supported me on the way towards my PhD. Specifically, I would like to thank my supervisor Prof. Dr. Frank Goldhammer and the co-authors of the three papers included in this dissertation: Dr. Ulf Kröhne and Prof. Dr. Johannes Naumann. Furthermore, I would like to thank my husband and son for their support.

Table of Contents

ZUSAMMENFASSUNG

Mit der rasanten Entwicklung neuer Technologien Ende des 20. Jahrhunderts haben vor allem solche Fähigkeiten an Wichtigkeit gewonnen, die sich auf den Umgang mit Informations- und Kommunikationstechnologien (engl. *information and communication technologies*, ICTs) beziehen. So hat beispielsweise der Aufbau und die rasche Weiterentwicklung des World Wide Web (Web) nicht nur grundlegend verändert, wie wir arbeiten, lernen und leben, auch unsere Fähigkeiten im Umgang damit sind maßgeblich für unseren schulischen Erfolg und unsere berufliche Karriere geworden. Als Informationsquelle bietet uns das Web mehr Möglichkeiten denn je. Dank Suchmaschinen wie Google, Yahoo! oder Bing und einer unbegrenzten Fülle an Informationen finden wir online eine Antwort auf alles. Doch sind wir auch in der Lage, die richtigen Antworten zu finden? Die vorliegende Dissertation befasst sich mit der Untersuchung zweier grundlegender ICT Fähigkeiten für die erfolgreiche Nutzung des Web als Informationsquelle: der Fähigkeit, auf die gewünschten Information zugreifen zu können, also basale Computerfähigkeiten (engl. *basic computer skills*, BCS), sowie der Fähigkeit, die Online-Informationen in Bezug auf ihre Glaubwürdigkeit bewerten zu können.

Hierzu wird zunächst der theoretische Hintergrund beider Konstrukte beschrieben sowie deren Relevanz dargelegt. Auf Basis der theoretischen Überlegungen werden Definitionen für die Konstrukte abgeleitet. Basale Computerfähigkeiten (BCS) werden definiert als *die fundamentale Fähigkeit und Geschwindigkeit, mit der basale Aktionen auf grafischen Nutzeroberflächen von Computern durchgeführt werden, um auf Informationen zuzugreifen, diese einzuholen und bereitzustellen*. Des Weiteren wird die Bewertung der Glaubwürdigkeit von Online-Informationen definiert als *die kognitive Fähigkeit, effizient Bewertungen über die Glaubwürdigkeit von Online-Informationen aus der Suchmaschinensuche vorzunehmen, indem Strukturmerkmale (engl. structural features) sowie Nachrichtenmerkmale (engl. message features) der Informationsquellen evaluiert werden*. Strukturmerkmale sind dabei solche, die sich auf die Darstellung der Online-Informationen beziehen und fundamental zum Aufbau der Ergebnisseite aus einer Suchmaschinenabfrage (engl. *search engine result page*, SERP) oder einer Webseite beitragen, während sich Nachrichtenmerkmale auf Inhalt und Gehalt der Online-Informationen beziehen. Es wird davon ausgegangen, dass kompetente Suchmaschinenbenutzer bei Glaubwürdigkeitsbeurteilungen von Online-Informationen (Links auf der Ergebnisseite einer Suchmaschinenabfrage sowie den dazugehörigen Webseiten) sowohl Struktur- als auch Nachrichtenmerkmale evaluieren. Dabei wird angenommen, dass die Qualität von

Strukturmerkmalen sowohl auf der Ergebnisseite einer Suchmaschinenabfrage als auch auf den dazugehörigen Webseiten zur Vorhersage der Qualität von Nachrichtenmerkmalen genutzt wird.

Die vorliegende Arbeit beschreibt die Entwicklung zwei neuer Testverfahren zur interaktiven, computer-basierten Erfassung basaler Computerfähigkeiten sowie der Fähigkeit zur Bewertung der Glaubwürdigkeit von Online-Informationen mittels Evaluation von Struktur- und Nachrichtenmerkmalen. Die faktorielle Struktur der BCS-Skala und des Tests zur Evaluation von Online-Informationen (engl. *Test for Evaluating Online Information*, TEO) sowie deren Beziehung verwandten kognitiven Konstrukten werden untersucht. Des Weiteren werden Ergebnisse aus der praktischen Anwendung des TEO genutzt, um zu einem tieferen Verständnis über den Suchprozess im Web (engl. *web search proccess*) zu gelangen und Faktoren zu identifizieren, die diesen beeinflussen. Dabei werden sowohl sogenannte Aufgabencharakteristika (engl. *task characteristics*) untersucht, die den Kontext des Suchprozesses bestimmen, als auch individuelle Prozesscharakteristika (engl. *individual process characteristics*), die durch die Person bestimmt sind, welche die Informationssuche vornimmt.

Um den Forschungszielen der Arbeit gerecht zu werden, wurden drei Studien durchgeführt, die zusammenfassend dargestellt werden. Die erste Studie befasst sich mit der Entwicklung der Skala zur Erfassung basaler Computerfähigkeiten. Sie wurde im Rahmen der nationalen Ergänzung im *Programme for International Student Assessment* (PISA) 2009 mit 15-jährigen Gymnasialschülerinnen und -Schülern durchgeführt. Da Fähigkeit und Geschwindigkeit wichtige Leistungsaspekte bei der Bearbeitung kognitiver Aufgaben darstellen, wurden beide zur Erklärung individueller Unterschiede im Antwortverhalten bei der Bearbeitung der BCS-Skala herangezogen. (1) Bezüglich der psychometrischen Eigenschaften der Skala wurde sowohl für BCS Fähigkeit als auch BCS Geschwindigkeit Eindimensionalität erwartet, da basale Computerfähigkeiten uns helfen, mit Aufgabenanforderungen umzugehen, die verschiedenen Software-Anwendungen gemein sind. Hinsichtlich des Zusammenhangs mit verwandten Konstrukten wurde untersucht, (2) inwieweit praktisches Computerwissen (engl. *practical computer knowledge*) sowie (3) die Fähigkeit und Geschwindigkeit zur Worterkennung (engl. *word recognition*) die BCS Leistung vorhersagen, (4) inwieweit eine Beziehung zwischen der objektiven BCS-Skala und einer Selbsteinschätzung der eignen Computerfähigkeiten besteht, (5) inwieweit der Einfluss des Geschlechts auf computerbezogene Fähigkeiten repliziert werden kann und (6) inwieweit BCS als zugrunde liegende Komponente elektronische Lesefähigkeit vorherzusagen vermag.

Die Hypothesen wurden anhand einer Stichprobe von 315 Gymnasialschülerinnen und –Schülern (50,5% weiblich, 42,5% männlich, 7% unbekannt, M = 15.87 Jahre) überprüft, denen jeweils die BCS-Skala und Aufgaben zur Erfassung des praktischen Computerwissens, zur Worterkennung sowie zum elektronischen Lesen vorgelegt wurden. Die Ergebnisse konnten die angenommene Eindimensionalität für BCS Fähigkeit und BCS Geschwindigkeit bestätigen. Praktisches Computerwissen und Worterkennung konnten als Prädiktoren von BCS identifiziert werden. Der Zusammenhang zwischen BCS und der Selbsteinschätzung der eigenen Computerfähigkeiten zeigte sich wie erwartet signifikant und gering für BCS Geschwindigkeit und die Selbsteinschätzung sowie moderat für BCS Fähigkeit und die Selbsteinschätzung. Männer zeigten eine bessere Leistung in BCS als Frauen, jedoch konnte durch das Geschlecht nur ein geringer Anteil der Varianz in BCS aufgeklärt werden. Elektronische Lesefähigkeit konnte durch die Prädiktoren BCS Fähigkeit und BCS Geschwindigkeit vorhergesagt werden.

Die zweite Studie behandelt die Entwicklung des Messinstruments zur interaktiven, computer-basierten Erfassung der Fähigkeit zur Bewertung der Glaubwürdigkeit von Online-Informationen. Des Weiteren exploriert die Studie die latente Struktur des Konstruktes sowie seinen Zusammenhang mit basalen Computerfähigkeiten, Worterkennung und logischem Denken (engl. *reasoning ability*). Es wurde angenommen, dass (1) der Bewertung der Glaubwürdigkeit von Online-Informationen basierend auf der Evaluation von Struktur- und Nachrichtenmerkmalen (engl. *evaluation of structural and messsage features*, ESMF), zwei trennbare Evaluationsfähigkeiten unterliegen, nämlich die Fähigkeit zur alleinigen Evaluation von Strukturmerkmalen (engl. *evaluation of structural features*, ESF) und die Fähigkeit zur alleinigen Evaluation von Nachrichtenmerkmalen (engl. *evaluation of message features*, EMF). (2) Da basale Computerfähigkeiten den Zugang zu Informationen im Web überhaupt erst ermöglichen, wurde angenommen, dass diese als Prädiktor die Bewertungskompetenz vorhersagen. (3) Als weiterer Prädiktor für die Bewertungskompetenz wurde Worterkennung untersucht, da die Bewertung von Online-Informationen grundsätzlich Lesefähigkeiten erfordert, insbesondere jedoch die Evaluation von Strukturmerkmalen basale Lesefähigkeiten auf Wortebene verlangt. Grundsätzlich beinhaltet die Bewertung der Glaubwürdigkeit von Online-Informationen Prozesse des Problemlösens, sodass erwartet wurde, dass (4) logisches Denken einen Effekt auf die Bewertungskompetenz hat. Um die Art des Zusammenhangs zwischen basalen Computerfähigkeiten, Worterkennung und logischem Denken mit der Bewertung der Glaubwürdigkeit von Online-Informationen zu spezifizieren, wurden weitere Hypothesen über direkte und indirekte Einflüsse dieser Kovariate auf ESMF, ESF und EMF formuliert. Anhand einer Stichprobe von 205 Gymnasialschülerinnen und –Schülern der 12. Klasse (42.4% weiblich, $M = 18.03$) wurden die Hypothesen mittels Daten, die im Rahmen einer Pilotstudie der *National Educational Panel Study* (NEPS) erhoben wurden, überprüft. Dazu wurden Testwerte aus den Aufgaben des TEO, jenen der BCS-Skala sowie aus Aufgaben zur Worterkennung und logischem Denken gemittelt zusammengefasst (engl. *item parcels*) und mit Hilfe latenter Regressionsanalysen untersucht. Die Ergebnisse bestätigen die Hypothese, dass ESF und EMF spezifische Evaluationsfähigkeiten darstellen, die jeweils eigene Anteile zu ESMF beitragen. Während basale Computerfähigkeiten sich wider Erwarten nicht als Prädiktor für die Bewertung der Glaubwürdigkeit von Online-Informationen erwiesen, konnten die Fähigkeit zur Worterkennung und jene zum logischen Denken jedoch als Voraussetzungen für die Bewertungskompetenz begründet werden. Als prädiktiv erwiesen sich Worterkennung und logisches Denken dabei insbesondere für die Evaluation von Strukturmerkmalen (ESF).

In der dritten Studie werden unter Anwendung des TEO der Suchprozess im Web mittels Suchmaschine näher beleuchtet und Bedingungen einer erfolgreichen Bewertung von Online-Informationen untersucht. Dabei wurde zunächst zwischen den zwei Einflussgrößen der Aufgabencharakteristika und individuellen Prozesscharakteristika unterschieden. Im Speziellen wurde der Einfluss dreier Aufgabencharakteristika geprüft, die sich auf die Komplexität einer Aufgabe beziehen: die Anzahl der Suchergebnisse (Links auf der Ergebnisseite einer Suchmaschinenabfrage), die Attraktivität der weniger glaubwürdigen Links auf der Ergebnisseite einer Suchmaschinenabfrage im Vergleich zum glaubwürdigsten Link sowie die Kongruenz zwischen den Glaubwürdigkeitsmerkmalen (Struktur- und Nachrichtenmerkmale) in den Links auf der Ergebnisseite der Suchmaschinenabfrage und auf den dazugehörigen Webseiten. Die Attraktivität eines Links wurde in diesem Zusammenhang definiert als die Gesamtanzahl der Struktur- und Nachrichtenmerkmale, welche auf eine hohe Glaubwürdigkeit der Informationen hinweisen. Je mehr Struktur- und Nachrichtenmerkmale eines Links demnach hohe Glaubwürdigkeit indizieren, desto attraktiver ist ein Link. Von Kongruenz zwischen den Glaubwürdigkeitsmerkmalen wurde ausgegangen, wenn die Struktur – und Nachrichtenmerkmale in einem Link auf der Ergebnisseite der Suchmaschinenabfrage und jene auf der korrespondierenden Webseite entweder gleichermaßen hohe oder gleichermaßen geringe Glaubwürdigkeit anzeigen. Inkongruenz hingegen wurde angenommen, wenn die Glaubwürdigkeitsmerkmale in einem Link auf der Ergebnisseite hohe Glaubwürdigkeit und jene auf der korrespondierenden Webseite geringe Glaubwürdigkeit anzeigen und vice versa. Weiterhin wurde der Einfluss dreier individueller

Prozesscharakteristika auf die erfolgreiche Bewertung der Glaubwürdigkeit von Online-Informationen untersucht: die Anzahl besuchter unterschiedlicher Webseiten bei der Informationssuche, die auf der Ergebnisseite der Suchmaschinenabfrage verbrachte Zeit sowie die auf den korrespondierenden Webseiten verbrachte Zeit (jeweils in Sekunden). Für die Attraktivität der weniger glaubwürdigen Links sowie die Anzahl der Links auf der Ergebnisseite wurde ein negativer Effekt auf die erfolgreiche Bewertung von Online-Informationen angenommen, während für die Kongruenz der Glaubwürdigkeitsmerkmale ein positiver Effekt erwartet wurde. Für die drei individuellen Prozessmerkmale wurde jeweils ein positiver Effekt auf die erfolgreiche Bewertung von Online-Informationen vermutet. Zudem wurden mögliche Interaktionen zwischen Aufgabencharakteristika und individuellen Prozesscharakteristika untersucht. Die Hypothesen wurden anhand einer Stichprobe von 379 Schülerinnen und Schülern (49% weiblich, $M = 16.69$ Jahre) der 9. Klasse aller Schulzweige anhand von Daten, die im Rahmen einer Pilotstudie der *National Educational Panel Study* (NEPS) erhoben wurden, untersucht. Diese bearbeiteten die Sub-Skala ESMF des TEO. Die Ergebnisse aus generalisierten gemischten linearen Modellen bestätigten den negativen Einfluss der Anzahl der Links sowie den positiven Einfluss kongruenter Glaubwürdigkeitsmerkmale. Es zeigt sich jedoch kein Effekt der Attraktivität der weniger glaubwürdigen Links auf den Aufgabenerfolg. Während die Anzahl besuchter unterschiedlicher Webseiten den Aufgabenerfolg positiv beeinflusste, zeigte sich kein Effekt für die beiden zeitbezogenen Prozesscharakteristika. Eine negative Interaktion zwischen der Anzahl der Links auf der Ergebnisseite und der Anzahl besuchter unterschiedlicher Webseiten implizierte, dass sich die Anzahl besuchter Webseiten bei einer hohen Linkanzahl auf der Ergebnisseite einer Suchmaschinenabfrage als weniger prädiktiv für den Aufgabenerfolg erweist als bei einer geringen Linkanzahl.

Im abschließenden Teil der Arbeit werden die Ergebnisse der drei vorgestellten Studien im Hinblick auf ihre Implikationen, Grenzen und zukünftige Forschungsfragen diskutiert. Die Arbeit trägt zu einem tieferen und differenzierten Verständnis wichtiger ICT-Fähigkeiten bei. Sie präsentiert reliable Instrumente zur Erfassung basaler Computerfähigkeiten sowie der Fähigkeit zur Bewertung der Glaubwürdigkeit von Online-Informationen, zeigt die hohe Relevanz basaler Lesefähigkeiten (Worterkennung) für beide Konstrukte sowie die Relevanz logischen Denkens für die Bewertungskompetenz. Die Betrachtung des Einflusses von Aufgabencharakteristika sowie individuellen Prozesscharakteristika auf die erfolgreiche Bewertung von Online-Informationen vertieft das Verständnis für die hinter der Bewertungskompetenz stehenden Prozesse.

1. Introduction and Overview

Starting with the advent of the World Wide Web (Web) in the 1990s and the associated rapid development of information and communication technology (ICT), computer-related skills have become increasingly important for our everyday life and have therefore been a significant research topic for some time now. In the recent decades, especially the competent and self-dependent handling of the great diversity of information found online has drawn researchers' attention (Gerjets, Kammerer, & Werner, 2011; Kammerer & Gerjets, 2014; Mason, Junyent, & Tornatora, 2014; Metzger, Flanagin, & Zwarum, 2003; Metzger, 2007; Pan et al., 2007). Through the Web more information is available than ever before (Metzger, 2007). The use of search engines such as Google, Bing or Yahoo! is quick, easy and cost-efficient and has therefore become a ubiquitous part of our personal and professional lives in order to fulfill our current information needs (Kammerer & Gerjets, 2014). But despite the obvious advantages of search engines, their use also poses a challenge to the user. In order to successfully make use of the Web as an unlimited source of information, we need to be ICT literate.

While there are many different approaches defining ICT literacy differing in their extent to which they focus on technological, informational or motivational aspects, the present dissertation regards ICT literacy as a context-specific cognitive ability that is obtained by the experience of dealing with specific ICT requirements and situations. According to the conceptualization of the International ICT Literacy Panel (2002) this cognitive ability can be separated into five constituent competencies increasing in complexity: "ICT Literacy is using digital technology, communication tools, and/or networks to access, manage, integrate, evaluate, and create information in order to function in a knowledge society" (p.2).

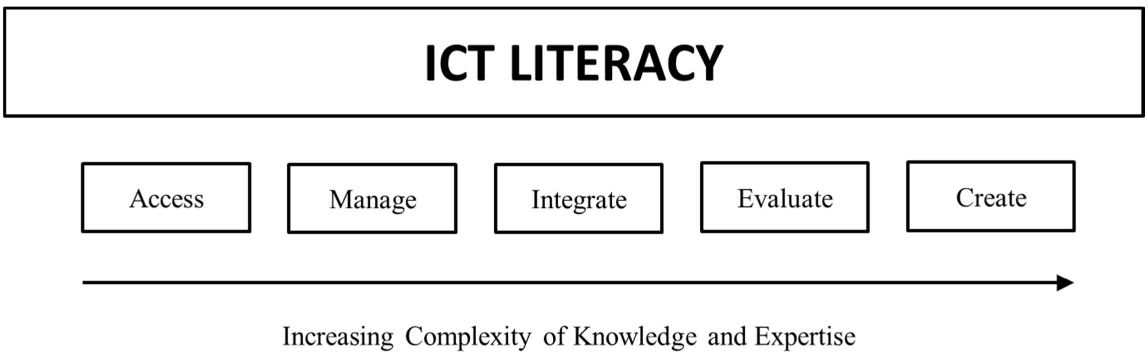


Figure 1. Components of ICT Literacy (International ICT Literacy Panel, 2002).

As the present dissertation is concerned with means of successfully using the Web as an information source, it explicitly addresses the competencies to *access* and *evaluate* information. In contrast to the competencies to manage, integrate and create information, they are highly specific for ICT. The ability to *access* information is anchored at the lower end of the continuum of ICT-related competencies and is a necessary precondition for the acquirement of the following competencies. It enables Web users to access the ICT environment to collect and retrieve information and therefore represents a particularly *technology-based* ability, which in the following will be referred to as *basic computer skills* (BCS). The ability to *evaluate* online information, that is “making judgements about the quality, usefulness or efficiency of information” (International ICT Literacy Panel, 2002, p. 3) is a higher-level competence of ICT literacy that is strongly *information-based*. It refers to the handling of a problem that has in this extent only evolved due to the introduction of the Web as an unlimited source of information highly differing with regard to its quality.

Models of ICT-related information processing like the IPS-I model (Brand-Gruwel, Wopereis, & Walraven, 2009) describe three layers of conditional skills that are of importance when performing an information problem solving (IPS) task in the ICT environment (e.g., the search for information on the Web), namely computer skills, evaluating skills and reading skills. Imagine for example a student that needs to find information for an essay or someone who wants to inform himself about different medical treatments because he has been diagnosed with a certain disease. In both cases Web users might start by using a search engine and it will be necessary to access, read, evaluate and integrate multiple information sources to solve the information problem. Thus, the ability to access information in an ICT environment involving basic technical computer skills as well as the ability to critically evaluate this information with respect to its’ credibility represent main components of ICT literacy and are necessary skills during the IPS process in an ICT environment such as successfully using Web-based information sources.

In the following chapters 2 and 3, an overview of the main constructs, basic computer skills and evaluating the credibility of online information, that are investigated in this dissertation, is given by illustrating their theoretical background and relevance. Subsequently, chapter 4 outlines the aim of the present dissertation, while chapters 5, 6 and 7 comprise summaries of the three studies that were conducted. A general discussion of the main findings, limitations and directions for future research are given in chapter 8.

2. Basic Computer Skills

This chapter described the construct of basic computer skills outlining its theoretical anchoring and relevance as well as presenting a definition for basic computer skills.

2.1 Theoretical Background

Given the increasing importance of ICT in peoples' everyday lives, computer-related skills have become "necessary to function effectively in a global economy increasingly dependent on ICT" (International ICT Literacy Panel, 2002, p. 6). To successfully make use of the Web as a rich information source, we first need the cognitive skill to access information, namely "knowing about and knowing how to collect and/or retrieve information" (International ICT Literacy Panel, 2002, p. 3). These basic ICT skills are not to be confounded with ICT, but can be seen as the lower-level part of ICT literacy. They are highly automatic and usually represented as procedural rather than declarative knowledge (see Anderson, 1982). The basic ICT skills refer to fundamental skills that can be considered core technical skills enabling one to perform simple actions being common to many software applications, e.g., clicking on a menu button or handling the file-management function (see Markauskaite, 2007). Such actions can be completed in a few steps and are usually encountered regardless of the current reason for using a particular software.

As *ability* and *speed* represent major aspects of human performance in cognitive tasks (for review see Carroll, 1993), considering both, together indicating efficiency, seems most appropriate when focusing on basic ICT skills reflected by relatively easy tasks (see the approach by Sandene, Bennett, Braswell, & Oranje, 2005, measuring both input speed and input accuracy as components of basic computer skills).

2.1.1 Definition of basic computer skills.

According to the preceding assumptions the construct of basic computer skills (BCS) is defined as *the fundamental ability and speed of performing basic actions in graphical user interfaces of computers to access, collect and provide information*.

The definition centers on information, i.e., ICT is considered as a tool applicable to dealing with information for solving tasks of daily life. The term computer is used in a general sense and covers the operation of desktop computers, which may be connected to networks, but also communication devices. The human computer interface is assumed to be graphical given the importance of visual perception in human information processing (e.g., Anderson, Matessa, & Lebiere, 1997). To represent programs and commands graphically, the user interface includes windows, icons, and menus that can be manipulated by a mouse and keyboard. Accessing information involves basic technical skills for making information available for acquisition and further processing, e.g., using links in a Web environment, navigation buttons of a Web browser, menus, and search functions. Collecting information is related to basic ways of gathering information to maintain accessibility, e.g., file management, creating bookmarks and connections, typing, as well as copying and pasting. Finally, providing information means information is made accessible for others, using basic technical functions for sharing information, e.g., by sending or forwarding e-mail messages in an e-mail client.

2.1.2 Relevance of the construct of basic computer skills.

The relevance of the construct of BCS is given in several aspects. First, BCS itself is a highly relevant competence construct. Nowadays, due to the growing specialization in all scientific areas individuals are confronted with a large number of problems for which the use of external information sources such as the Web is indispensable (Bromme, 2005; Rouet et al., 2009). But whether individuals use the Web in schools and universities or in more informal learning contexts, they need sufficient basic computer skills. Individuals lacking these skills will clearly be left behind both in educational as well as in professional contexts (see Leu, Kinzer, Coiro, & Cammack, 2004). Related research requires measures of ICT literacy to assess it cross-sectionally or longitudinally, e.g., to study the effect of instructional interventions aiming at the improvement of ICT literacy or to investigate the development of ICT literacy across the lifespan (e.g., Weinert et al., 2011). Second, the BCS construct has become an important control variable as many assessments increasingly rely on computer based assessment procedures. But if the individual ability to operate a computer affects the test score above and beyond the cognitive disposition to be assessed, the validity of the test score is threatened (e.g., Parshall, Spray, Kalohn, & Davey, 2002; Russell, Goldberg, & O'Connor, 2003; Sandene et al., 2005). In consequence, when developing computer-based assessments of cognitive dispositions, it is methodologically crucial to minimize the influence of individual differences in computer-related skills or control by assessing them as covariate. Third, the computer itself has become an important means of instruction, and materials to be learned are delivered through the computer as hypertext, hypermedia, or simulations. In research on learning with the computer, ICT skills are generally an important covariate to be measured (e.g., Wecker, Kohnle, & Fischer, 2007).

3. Evaluating the Credibility of Online Information

This chapter is concerned with the construct of evaluating the credibility of online information, its theoretical background and relevance. It outlines the characteristics of the Web search process and the cognitive processes involved. Furthermore, criteria affecting credibility judgements during the Web search process are described. The chapter concludes with a definition of the construct of evaluating the credibility of online information based on the preceding statements.

3.1 Theoretical Background

The development and rise of the Web has certainly changed the way we learn, work and live. Nowadays the use of search engines to fulfill our current information needs can be considered part of our everyday lives (Kammerer & Gerjets, 2014). Going online we can find an answer to everything including questions we had not even asked. But are we able to find the right answers? Due to the unlimited amount of information that mostly does not underlie any editorial review, the ability to find credible information has become and essential skill (Hargittai, Fullerton, Menchen-Trevino, & Thomas, 2010; Hilligoss & Rieh, 2008; Metzger, 2007; Rieh, 2002; Schwarz & Morris, 2011). However, research has shown that individuals differ substantially in their ability to evaluate online information with respect to its credibility and even students, being frequent users of the Web, struggle in differentiating less credible from reliable Web sites (Lorenzen, 2001; MaKinster, Beghetto, & Plucker, 2002; Wallace, Kupperman, Krajcik, & Soloway, 2000; Walraven, Brand-Gruwel, & Boshuizen, 2009). But it has also been proven that the ability can be improved by adequate training programs. Therefore, research requires a measurement tool, not only to assess an individuals' level of performance and identify the individual need for practice but also to study the effect of instructional interventions and investigate the development of the ability across the life span (e.g., Weinert et al., 2011). Furthermore, especially in current times with the rise of so-called "fake news" (see Antos, 2017), deliberate misinformation or hoaxes spread on the Web particularly in social networks, the ability to evaluate the credibility of online information is a highly relevant competence construct that yet has only been explored in parts. To gain deeper insights into this complex construct and the specific processes behind, it is important to further examine influencing variables. Findings on which characteristics determine a successful information evaluation on the Web might then in turn foster the development of adequate training programs for students.

3.1.1 The web search process.

Evaluating the credibility of online information is part of the Web search process, a multi-staged process describing how Web users search for information on the Web and solve information processing problems online. The Web search process has been described by several models and is represented by five important stages (Gerjets et al., 2011, p. 221): (a) An information deficiency is identified and a search goal is defined. (b) To access the information in demand, a search engine (e.g., Google) is selected, a search terms is entered and the search is started. In the following, a so called ''search engine result page'' (SERP) with a list of search result links is returned to the user. (c) The list of search results is scanned and the links are evaluated with regard to their significance for the search goal. Links are selected for further inspection. At this time the evaluation of search results is based on sparse information about the corresponding Web sites and the information they contain (e.g., titles, text fragments, domain names and suffixes). One major (observable) action in the information search process is the clicking on a link and entering the corresponding Web site. (d) After entering a selected Web site, its information is scanned, evaluated with regard to its relevance for the search goal, and in case of relevance information is extracted for further processing. (e) The information from different Web sites is compared, evaluated with regard to its credibility and integrated towards a solution of the information problem.

The construct of evaluating the credibility of online information as conceptualized in paper 2 and 3 of the present dissertation explicitly addresses the evaluation of links on SERPs and corresponding Web sites as well as the comparison of the different links and Web sites with respect to their credibility. Therefore, the focus of these papers is on the last three stages of the Web search process because these are the stages in which the Web user needs to evaluate online content in contrast to the first two stages, which demand the Web user to access online information and therefore different cognitive skills are required.

In the following the properties of a Web site are elaborated, which are assumed to reflect the credibility of online information, and, therefore affect individual credibility judgments when evaluating online information. Furthermore, two major factors potentially influencing a successful information evaluation online are described in greater detail: the context, in which the web search task takes place, defined by various characteristics (the so-called "task characteristics"), and the individual performing the Web search task by applying information processing strategies indicated by behavioral process characteristics.

3.1.2 Criteria affecting credibility judgements during web search.

In the following it is elaborated on different criteria affecting credibility judgements during Web search. Information features, task characteristics and individual process characteristics are investigated and described in further detail.

3.1.2.1 Information credibility features.

With the rise of user-generated content on the Web and the urgency to properly evaluate online information, the question of which skills are needed to successfully cope with this task has been topic of research for some time now and produced several models explaining information evaluation behavior online. According to Prominence-Interpretation-Theory (P-I-Theory, Fogg, 2002) Web users first notice an information feature (Prominence) and then make a judgment about it (Interpretation). Both steps are necessary conditions for a credibility assessment. This process of noticing and interpreting information features will typically happen numerous times during the evaluation of online information, with new aspects on a SERP or Web site being noticed and interpreted (Fogg, 2002). Thus, evaluating online information refers to the judgment of many different information characteristics, which can be classified as features reflecting the content of information and features reflecting its presentation (Lucassen & Schraagen, 2011). Similarly Wathen and Burkell (2002) propose a model including two categories when evaluating online information: evaluation of surface credibility based on surface characteristics and evaluation of message credibility.

Prior research has repeatedly identified five quality attributes that need to be considered when judging the credibility of print and online information found in factual texts, which can be considered to be true or false: accuracy, authority, objectivity, currency, and coverage or scope (Alexander & Tate, 1999; Brandt, 1996; Fritch & Cromwell, 2001; Kapoun, 1998; Meola, 2004; Scholz-Crane, 1998; Smith, 1997). Metzger argues that “Skills needed to determine the quality or credibility of online information are largely the same as those for evaluating information found in other channels of communication” (Metzger, 2007, p. 2079). But clearly recent research on online content and credibility includes information features that did not exist during the time of early research on information credibility (Hong, 2006). Therefore, the examination of online information requires additional skills beyond those required for the evaluation of offline information (Castek, Zawilinski, McVerry, O'Byrne, & Leu, 2011; Coiro, Knobel, Lankshear, & Leu, 2008). Quality attributes unique to the Web which cannot be found in print information, e.g., navigation tools such as site maps or domain names and suffixes need to be considered when searching credible information on the Web. Most important, these attributes unique to online information are assumed to be predictive for the credibility of online search results. They concern the presentation of the online information and "can be called structural features because they constitute elements fundamental to the composition of Web sites" (Hong, 2006, p. 115). Thus, in order to evaluate important information characteristics the information searcher needs to investigate message features reading the text and/or investigate accompanying information relating to its presentation (structural features). In the remainder of this dissertation we refer to the *evaluation of message features* (EMF) when investigating the message content by reading a text, while we refer to the *evaluation of structural features* (ESF) when investigating accompanying information found in its presentation.

As for message features and credibility, an objectively verifiable relationship is also assumed for structural features and credibility. Based on this assumption, a Web site is assumed to be a credible source of information if the message features concerning text quality are fulfilled, whereas the presumed structural features for high quality Web sites allow this conclusion in advance with high probability. Thus, both, message and structural features indicate the overall credibility of a Web site and therefore both need to be considered when searching and evaluating online information (Hong, 2006).

3.1.2.2 Task characteristics.

As it has become increasingly important to understand in detail how Web users evaluate online information, identifying factors influencing decision-making processes during Web search provides a useful approach. Tasks characteristics, determining the context in which the Web search takes place, are supposed to invoke specific search processes and in consequence influence an individual's success when evaluating online search results (Wirth, Böcking, Karnowski, & von Pape, 2007). Dual-processing models such as the Heuristic Systematic Model (HSM) or the Elaboration Likelihood Model (ELM) (Chen & Chaiken, 1999; Petty & Cacioppo, 1986) describe two modes of information processing: systematic processing via a central route and heuristic processing via a peripheral route. Whereas systematic information processing implies a complete evaluation of all given information (e.g., message features such as currency, completeness, objectivity etc.) and therefore requires a larger amount of cognitive resources, heuristic information processing relies on superficial cues (e.g., structural features such as domain name, page rank etc.) that are expected to require less cognitive resources. Which mode of information processing is carried out, is amongst other factors such as cognitive ability moderated by task characteristics (Chaiken, Liberman, & Eagly, 1989; Evans, 2008; Petty & Cacioppo, 1990). Several task characteristics influencing the mode of information processing during Web search have been studied in the past: on a macro level task type and task complexity (Wirth et al., 2007), and on a micro level the position of the search result on a SERP (Pan et al., 2007; Salmerón, Kammerer, & García-Carríon, 2013), the number of links returned by the search engine and their characteristics, the complexity and structure of a web site etc. (Wirth et al., 2007).

To conclude, task characteristics in general encourage specific ways of information processing during Web search, which might rise or reduce the probability of a successful information evaluation. For instance, although the application of heuristic processing strategies are widespread and in many cases sufficient (Wirth et al, 2007) some task characteristics might encourage Web users to use heuristics leading to an insufficient information evaluation. A task characteristic such as a high number of links returned by the search engine, might motivate the Web user to minimize cognitive effort and rely on the so called “top-link heuristic” (Pan et al., 2007) choosing the top link of a SERP in believe that the top links are often the best ones neglecting the fact that even the top search results might not be credible information sources and one-sided or commercially biased due to search engine optimization business (Lewandowski, 2011, 2013).

The present dissertation is concerned with the investigation of three task characteristics reflecting task complexity: the number of information sources that Web users need to evaluate during Web search, the degree of attractiveness of the different information sources that need to be compared and the congruency between credibility features (structural and message features) present on the SERP and those present on the corresponding Web site. When the number of information sources to be compared and evaluated is low (i.e., task complexity is low), Web users have more cognitive resources available, e.g., working memory (see Baddeley, 2012), facilitating a systematic and deep information processing. In consequence Web users are expected to invest more time on evaluating the different information source resulting in a higher probability of successfully identifying the most credible source.

The attractiveness of an information sources to be evaluated is defined by the total amount of credibility features (structural and message features) indicating high credibility of the source. Consequently, the more credibility features indicate high credibility, the higher the attractiveness of an information source. The higher the overall attractiveness of the less credible information sources (i.e., task complexity is high), the more difficult it should be to identify the most credible source.

The congruency between credibility features (structural and message features) present on the SERP and those present on the corresponding Web site is also expected to influence a successful information evaluation. The credibility features on a SERP and the corresponding Web site can either be congruent both indicating high or low credibility or be incongruent with the credibility features on the SERP indicating high information credibility, while those on the corresponding Web site indicate low information credibility and vice versa. As dealing with conflicting, controversial or multi-perspective information requires a large amount of cognitive resources such as a heightened level of concentration and WM capacity (e.g., Kornmann et al., 2016) incongruent credibility features on the link of a SERP and the corresponding Web site should result in high task complexity. Thus, it should be easier to successfully identify the most credible information source when credibility features of a link on the SERP and of the Web site are congruent (i.e., task complexity is low) because no conflicting information needs to be processed and more cognitive resources are available for evaluation processes.

3.1.2.3 Individual process characteristics.

In addition to task characteristics, a successful information evaluation online is influenced by characteristics of the individual carrying out of the Web search. As individuals differ in their ability to successfully evaluate online information, they can be expected to carry out individually different processing strategies while conducting a Web search. Thus, it is assumed that competent Web searchers are likely to show a specific Web search behavior that can be differentiated from the evaluation processes carried out by less competent Web searchers. In order to explore search and evaluation processes further empirically, traces of the Web search process have to be identified by investigating process data. Subsequently these traces of processing can be used to derive indicators providing important information about the individual search processes and in consequence terms of successful information evaluation online. Several observable actions reflecting the process of Web search can be used to compose these process indicators, for example clicking on a link on a SERP and accessing the corresponding Web site. Beside these observable actions time-related information is an important indicator providing information about individual differences in the Web search process. In this work, the number of visited Web sites, the total time spent on the SERP and the total time spent on the different Web sites visited were investigated with respect to their influence on a successful information evaluation online.

3.1.3 Definition of evaluating online information.

Based on the preceding statements, the construct of evaluating the credibility of online information (evaluating structural and message features, ESMF) is defined as *the cognitive ability to efficiently make judgments about the credibility of online information accessed by a search engine by taking structural and message features into account*. A competent Web searcher should bear structural and message features in mind while evaluating results on a SERP and the information on corresponding Web sites. Structural features are assumed to be used as predictors for message features both on a SERP and on a selected Web site.

4. Aim of the Dissertation

The overall goal of this dissertation is to investigate the constructs of basic computer skills (BCS) and evaluating the credibility of online information using structural and message features (ESMF) from an individual differences perspective. Newly developed scales for the assessment of both ICT skills utilizing interactive computer-simulated tasks are introduced. The factor structure of the constructs and their relation with related constructs are investigated. Furthermore, applying the newly developed measurement tool for assessing the ability to evaluate online information, the dissertation aims at gaining a deeper understanding of the Web search process by investigating influencing variables. Tasks characteristics and individual process characteristics during Web search activities are explored as potential predictors for a successful information evaluation online. The three main goals of this dissertation and related hypotheses are described in detail in the following.

1. Development of a measurement tool for BCS.

The first main goal of this dissertation is the development of an interactive, computer-based scale assessing basic computer skills with tasks simulating typical computer environments. Psychometric properties of the newly developed BCS scale as well as its relation with related constructs are investigated by addressing the following hypotheses:

(1) Unidimensionality is assumed for BCS speed and BCS ability since BCS enable us to deal with task requirements that are common to many software applications and purposes. BCS speed and BCS ability are assumed to be positively related dimensions.

(2) As domain knowledge is assumed to be a necessary condition for competent behavior (see Mayer, 2003), practical computer knowledge is assumed to be strongly associated with BCS. Furthermore, as BCS tasks include detecting and reading simple verbal labels, lower level reading skills, specifically ability and speed in word recognition are assumed to be associated with BCS performance. Effect sizes for word recognition are assumed to be smaller than those for practical computer knowledge.

(3) BCS ability and speed are assumed to be positively related to self-reported computer skills. Correlations are assumed to be moderate, at most .30.

(4) As prior research has indicated, that male students have better ICT skills than females (e.g., Ilomäki & Ratanen, 2007; Kuhlemeier & Hemker, 2007), male participants are assumed to show higher mean levels of BCS ability and speed than females.

(5) BCS speed and ability are assumed to be predictors of electronic reading ability as electronic reading requires the handling of computer interface (see Leu et al, 2004; OECD, 2011).

2. *Development of a measurement tool for ESMF and exploration of selected cognitive covariates to the ability of ESMF.*

The second goal of this dissertation is the development of an interactive, computer-based measurement tool assessing the ability to evaluate the credibility of online information. Furthermore, to shed some light on the complex construct of ESMF, its latent structure and relationship with selected cognitive components is explored.

(1) The construct of ESMF, incorporating the evaluation of message and structural features, underlies the evaluation of structural features only (ESF) and the evaluation of message features only (EMF). ESF and EMF are assumed to contribute uniquely to ESMF.

(2) Basic computer skills are assumed to explain the evaluation of online information and have a stronger direct effect on ESMF than on its constituents ESF and EMF due to varying hypertext navigation demands. The direct effect of basic computer skills on ESMF is also assumed to be stronger than possible indirect effects of basic computer skills on ESMF via ESF or EMF.

(3) As ESF includes detecting and reading structural features, e.g., domain names and suffixes, a stronger direct effect of word recognition on ESF than on ESMF and EMF is assumed. The indirect effect of word recognition via ESF on ESMF is assumed to be stronger than the indirect effect via EMF on ESMF.

(4) Reasoning ability is assumed to predict ESF because ESF involves the prediction of message features from structural features based on heuristic rules. The direct effect of reasoning ability on ESMF and EMF is not assumed to be as strong as the direct effect on ESF because both allow a direct judgment of message features and do not necessarily require making inferences. The indirect effect of reasoning ability via ESF on ESMF is assumed to be stronger than the indirect effect via EMF on ESMF.

3. *Investigation of the influence of task characteristics and information processing behavior on the successful evaluation of online information.*

The third goal of the dissertation is to gain deeper insights into the Web search process by investigating the influence of task characteristics and individual process characteristics on a successful information evaluation online. Task characteristics provoking task specific search processes are identified by investigating log-file-data and assumed to affect task success. They can be assumed to reflect the Web search and evaluation process and affect task success in turn. Moreover, individual process characteristics that serve as predictors for a successful evaluation of online information are identified. The first three hypotheses address the influence of task characteristics on a successful information evaluation online.

(1) High attractiveness of non-target search results has a negative effect on the successful task completion, that is, the task is more difficult.

(2) When internet users are confronted with a high number of information sources, this is expected to have a negative effect on the probability of task success because the task is more difficult.

(3) Congruency between credibility features of the SERP and linked web pages has a positive effect on the probability of task success, that is, the task is expected to be less difficult.

The following three hypotheses refer to the individual process variables assumed to have an impact on a successful evaluation of online information.

(4) The number of visited Web sites offered on a SERP is assumed to have a positive effect on the successful completion of the task. In case more Web sites are visited, the higher the probability of task success.

(5) The time spent on a SERP is assumed to have an effect on successful task completion. The more time is spent on a SERP, the higher the probability of task success.

(6) The time spent on the different web sites linked on a SERP is assumed to have an effect on successful item completion. The more time is spent on the different Web sites, the higher the probability of a thorough, in-depth and complete processing of all given information, which is more likely to result in task success.

In the following, the three studies of this dissertation are briefly presented to investigate the hypotheses raised.

5. Summary of Paper 1: Assessing Individual Differences in Basic Computer Skills: Psychometric Characteristics of an Interactive Performance Measure

In this paper the development of a scale for the assessment of basic computer skills (BCS) utilizing interactive simulations is described. The BCS scale was developed for the national extension of the Program for International Student Assessment (PISA) 2009 study targeting German 15-year-old secondary school students. The tasks of the BCS scale were designed to simulate typical computer environments and require responses within these environments using mouse and/or keyboard.

The first hypothesis addressed the psychometric properties of the newly developed BCS scale. (1) Since BCS enables us to deal with task requirements that are common to many software applications, unidimensionality was assumed in both BCS speed and BCS ability. The following hypotheses referred to the relation of the BCS scale with related constructs. (2.1a) For BCS it was assumed that practical computer knowledge, i.e., the knowledge about how to solve everyday computer problems, facilitates the development of BCS. The relationship of BCS and computer knowledge was assumed to be bidirectional: While good knowledge helps one to develop good BCS, good BCS also helps in the acquisition of new knowledge procedures. As a consequence it was assumed that practical computer knowledge has strong associations with BCS ability and BCS speed. (2.1b) Ability and speed in word recognition were expected to be associated with BCS performance. (2.1c) Taking both practical computer knowledge and word recognition into account, effect sizes for word recognition were assumed to be smaller than those for practical computer knowledge. (2.2) BCS ability and speed were assumed to be positively related to self-reported computer skills. Correlations were expected to be moderate, that is, at most about .30. (2.3) Male participants were expected to show higher mean levels of BCS ability and speed than females. (2.4) BCS ability and speed were assumed to be predictors of electronic reading ability.

Participants were 315 secondary school students (50,5% females, 42,5% males, 7% unknown) with an overall mean age of 15.87 (SD = .28) ranged from 15.42 to 16.33 years. The following variables were measured: BCS was assessed applying the newly developed Basic Computer Skills Test comprising 15 tasks with individual log-transformed response time (RT) and the response (R) being collected for each task. Computer knowledge was assessed with the scale *practical computer knowledge* (PRACOWI) comprising 20 tasks from a German inventory for the assessment of computer literacy, computer-related attitudes and computer anxiety (Revised Computer Literacy Inventory, INCOBI-R, Richter, Naumann, &

Horz, 2010). Speed and ability of word recognition (WR) were assessed with a Lexical Decision Task (see, e.g., Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004). The self-report on computer skills included 12 tasks (for more details see OECD, 2011, chapter 5) and electronic reading ability was assessed by 67 text comprehension tasks in 13 short hypertexts (see OECD, 2011, chapters 2 and 3). Students were tested in groups of 5-10 in computer labs at schools. The overall completion time was about 120 minutes.

Psychometric properties of the BCS scale were analyzed using confirmatory factor analysis (CFA) (e.g., Bollen, 1989) and by considering coefficients related to classical test theory (CTT). CFAs were conducted to test the hypothesized unidimensional measurement models for BCS ability and speed and to assess the indicators' properties and to explore the covariance structure assumed to exist between BCS ability and speed. Model parameters were estimated by means of the Mplus software (Muthén & Muthén, 1998-2011). A model was considered to show a good/acceptable overall model fit if the model fit criteria according to Muthén (1998-2004) and Schermelleh-Engel, Moosbrugger, & Müller (2003) were met. CTT-related item coefficients and Cronbach's α were estimated using statistical packages of the R environment (R Development Core Team, 2011). For investigating the validity of the BCS scale, Mplus was used to estimate regression and correlation coefficients.

Results revealed unidimensionality of BCS speed and ability after dropping two ill-fitting tasks, which supported the assumption that BCS seem to represent generic skills not being specific to particular environments. BCS ability and BCS speed were also substantially related to one another, sharing about one fourth of their variance, indicating that accurate participants also tend to be fast. However, the amount of unique variance led to the conclusion that both are specific constituents of BCS. Practical computer knowledge showed strong associations with BCS speed (*b=.47, p<.01*) and BCS ability (*b=.60, p<.01*). Furthermore, WR speed was shown to predict BCS speed (*b=.36, p<.01*) and WR ability was revealed to predict BCS ability (*b=.32, p<.01*), which lead to the conclusion that basic reading is a prerequisite for solving basic ICT tasks. The multiple regression of BCS speed on practical computer knowledge and WR revealed significant associations of BCS speed with practical computer knowledge (*b=.40, p<.01*), and with WR speed (*b=.30, p<.01*) when accounting for the respective other predictor. Regressing BCS ability on practical computer knowledge and WR ability, again significant associations were obtained for practical computer knowledge (*b=.55, p<.01*), and for WR ability (*b=.18, p<.01*). As expected the correlation between self-reported computer skills and BCS speed was small but significant (*r=.18, p=.04*) as well as the correlation of self-reported computer skills with BCS ability, which was moderate and significant (*r=.26, p<.01*). Results also revealed higher mean levels of BCS ability and speed for males. However, the amount of explained variance by gender was only very small (R^2*=.04* for BCS speed and R^2*=.06* for BCS ability). Finally, electronic reading was predicted by BCS ability (*b=.45, p<.01*) and BCS speed (*b=.21, p=.03*). According to the relations found with other measures and the results obtained, it was concluded that the newly developed BCS scale represents a reliable and valid measurement tool.

6. Summary of Paper 2: Evaluating the Credibility of Online Information: The Influence of Basic Computer Skills, Word recognition and Reasoning

This paper is concerned with the development of an interactive, computer-based test for the assessment of the ability to evaluate the credibility of online information. Furthermore, it investigated the evaluation of online information based on structural and message features (Evaluation of Structural and Message Features, ESMF), that was assumed to underlie two separable evaluation skills, namely the evaluation of online search results based on structural features only (Evaluation of Structural Features, ESF) and the evaluation of online search results based on message features only (Evaluation of Message Features, EMF) and explored its relation with basic computer skills, word recognition and reasoning ability. The first hypothesis addressed the constituent skills of ESMF, while the following hypotheses referred to the underlying conditional cognitive skills.

(1) ESF and EMF were assumed to represent specific evaluation skills of ESMF and contribute uniquely to ESMF. (2) Basic computer skills were assumed to be a contributory factor for the evaluation of the credibility of online information (Tu, Shih, & Tsai, 2008). Basic computer skills, i.e., having a notion of a hypertext structure and being able to navigate in an online environment to access information etc., were assumed to have a stronger direct effect on ESMF than on ESF and EMF due to varying hypertext navigation demands. The direct effect of basic computer skills on ESMF was also assumed to be stronger than possible indirect effects of basic computer skills on ESMF. (3) Lower level reading skills as word recognition were proposed to predict and underlie the evaluation of online information (Brand-Gruwel et al., 2009; Salmerón, Naumann, García, & Farjado, 2017). As ESF includes detecting and reading structural features, e.g. domain names and suffixes, a stronger direct effect of word recognition on ESF than on ESMF and EMF was proposed. The indirect effect of word recognition via ESF on ESMF was assumed to be stronger than the indirect effect via EMF on ESMF. (4) As the evaluation of online information involves processes of problem solving, it was expected to be predicted by reasoning ability (Funke & Frensch, 2007). More specifically, reasoning ability was assumed to predict ESF because ESF involves the prediction of message features from structural features based on heuristic rules. The direct effect of reasoning ability on ESMF and EMF was not assumed to be as strong as the direct effect on ESF because both allow a direct judgment of message features and do not necessarily require making inferences. Consequently, the indirect effect of reasoning ability via ESF on ESMF was expected to be stronger than the indirect effect via EMF on ESMF.

Participants were 205 high school students in grade 12 (42.4% female, 57.6% male) at nine schools in Germany, with an overall mean age of 18.03 (SD = .70) ranged from 15.5 to 21.75 years. The following variables were measured: The newly developed Test for Evaluating Online Information (TEO) consisting of 24 interactive items assigned to three subscales was applied to assess the cognitive ability to efficiently make judgments about the credibility of online information using structural and message features. Basic computer skills were assessed applying a revised and complemented version of the Basic Computer Skills test by Goldhammer et al. (2012) comprising 20 interactive tasks. Word recognition was assessed with a Lexical Decision Task (Balota et al., 2004; Richter, Naumann, Isberner, & Kutzner, 2011) asking students to judge eight words and eight non-words (16 items) appearing successively on a computer screen Finally, reasoning ability was assessed with 16 items measuring fluid intelligence from the Berlin Test of Fluid and Crystallized Intelligence (BEFKI, Schipolowski, Schroeders, & Wilhelm, 2008) requiring participants to complete rows of geometric designs based on heuristic rules to be induced.

Data was collected in the context of a pilot study of the National Educational Panel study (NEPS) sponsored by the Federal Ministry of Education and Research in Germany. The mentioned tests as well as two additional tests that were not included in the present study were administered in groups of up to 24 students in computer labs at schools. The overall completion time was about 120 minutes with a 5 minute break after 60 minutes.

Data analyses were based on both dichotomous item responses and item parcels. To determine psychometric properties of the TEO subscales and item indicators' properties Confirmatory Factory Analysis (CFA) for categorical data was used and item coefficients difficulty and discrimination related to classical test theory (CTT) were examined. CFA/SEM parcels as indicators were used to investigate the association of basic computer skills, word recognition and reasoning ability with ESMF, ESF and EMF. Model parameters were estimated by means of the Mplus software version 6 (Muthén & Muthén, 1998-2011). Item coefficients related to CTT were estimated using the ltem package (Rizopoulos, 2006) of the R environment (R Development Core Team, 2011). A model was considered to show a good/acceptable overall model fit if the model fit criteria according to Muthén (1998-2004) and Schermelleh-Engel et al., (2003) were met. To test if two nested models differ significantly, the Wald Test was used (Bollen, 1989). For the statistical tests of significance an alpha level of .05 was assumed.

Results proposed that ESF and EMF contribute uniquely to ESMF and represent specific evaluation skills, indicating that the evaluation of online information underlies different skills based on which features of online information are paid attention to and are evaluated with respect to their credibility. ESF and EMF were substantially related to ESMF explaining 76% of the variance of ESMF. In contrast to the expectations, basic computer skills did not reveal to be a significant predictor for ESMF and its constituents ESF and EMF. However, a significant direct effect for word recognition on ESF ($\beta=.34$, $p<.01$) was found. While there was no significant direct effect for word recognition on ESMF and EMF, path analysis revealed a significant indirect effect of word recognition on ESMF via ESF ($\beta=.29$, $p<.05$). In line with previous research (Brand-Gruwel et a., 2009), the results stressed the assumption that the strong effect of word recognition on ESF can be explained by the underlying process of scanning and comparing of structural features, which specifically involves verbal decoding and the recognition of words and meaningful letter sequences. As expected, results showed a significant direct effect of reasoning ability on ESF ($\beta=.51$, $p<.01$) and EMF ($\beta=.30$, $p<.05$), whereas there was no significant direct effect of reasoning ability on ESMF. It was speculated, that the missing direct effect of reasoning ability on ESMF might be due to the design of the ESMF subscale as half of the items contained incongruent structural and message features, which makes drawing inferences impossible. However, re-analyses of the data without the four incongruent items showed that the effect of reasoning ability on ESMF remained insignificant Furthermore, in line with the expectations, a significant indirect effect of reasoning ability on ESMF via ESF ($\beta=.44$, $p<.05$) was found, while there was no significant indirect effect of reasoning ability on ESMF via EMF. Overall, the study showed that word recognition and reasoning ability are important prerequisite skills for Web based information evaluation. Therefore, trainings for Web users might benefit from including the improvement of basic cognitive skills. This particularly refers to cognitive skills that can be modified by learning and instruction such as basic reading skills.

7. Summary of Paper 3: Successful Information Evaluation Online: The Role of Task Characteristics and Information Processing Behavior

This paper focuses on the Web search process and means of successfully evaluating the credibility of online information. Nowadays in need for information most people rely on the Web as their predominant information source, often basing important personal or job-related decisions on the information they find online. Therefore, is has become increasingly important to understand how people make use of search engines and which variables determine a successful information evaluation on the Web. Thus, this paper investigated two major factors for successful information evaluation, namely being able to identify the most credible search result on a search result page (SERP): the context in which the Web search takes place and the individual performing the Web search using information processing strategies. The first was defined by three task characteristics, which were (a) the number of information sources that need to be compared, (b) the attractiveness of the less credible information sources and (c) the congruency between credibility features present on the SERP and the corresponding Web site, while the latter was indicated by behavioral process characteristics such as (d) the number of visited Web sites, (e) the time spent on SERPS and (f) the time spent on Web sites.

While the first three hypotheses addressed the predictive validity of task characteristics determining task difficulty, the following three hypotheses referred to the influence of individual process characteristics on task success. (1) High-attractiveness of non-target search results was expected to have a negative effect on task success, that is, the task is more difficult. (2) A high number of information sources that need to be evaluated were assumed to have a negative effect on task success, again, the task is more difficult. (3) Congruency between credibility features in the links on the SERP and the corresponding Web sites was expected to have a positive effect on task success, that is, the task is expected to be less difficult. (4) It was assumed that the more Web sites are visited, the higher the probability of task success. (5) It was expected that the more time is spent on a SERP, the higher is the probability of task success. (6) And last, the more time is spent on the different Web sites, the higher was assumed the probability of task success.

Participants were 379 students in grade 9 from all three German school tracks (49% female, 51% male) at schools in Germany, with an overall mean age of 16.69 (SD = .67) ranged from 15 to 19 years. The cognitive ability to efficiently make judgments about the credibility of online information was assessed by the subscale Evaluating Structural and

Message Features (ESMF) from the Test for Evaluating Online Information (TEO) (see Keßel, Goldhammer & Kröhne, under review) consisting of eight interactive items simulating Web environments requiring the test taker to identify the most credible search result from a SERP and the corresponding Web sites.

Data was collected in the context of a pilot study of the National Educational Panel study (NEPS) sponsored by the Federal Ministry of Education and Research in Germany. The TEO and additional tests that were not included in the present study were administered in groups of up to 24 students in computer labs at schools. The overall completion time was about 120 minutes with a 15 minute break after 60 minutes.

Data analyses were based on dichotomous item responses and log-file data such as log-transformed reaction times (measured in seconds) and frequencies of Web site visits. To investigate, how task process characteristics and individual process characteristics determine a successful information evaluation, regression analyses within the Generalized Linear Mixed Modelling (GLMM) framework were conducted by means of the lme4-package (Bates, Maechler, Martin, & Bolker, 2011) of the R environment (R Development Core Team, 2011). Possible interactions between task characteristics and individual process characteristics were explored. To test if two models differ significantly, a model difference test (ANOVA) was conducted. For the statistical tests of significance an alpha level of .05 was assumed.

Concerning the task characteristics the attractiveness of non-targets did not show a significant effect on task success. In contrast the number of information sources had a significant negative effect on task success ($b=-.54, p<.05$) and the congruency of credibility features showed a significant positive effect on task success ($b=.62, p<.05$). For the individual process characteristics it was found that the number of visited Web sites had a highly significant positive effect on task success ($b=.44, p<.001$), while neither the time spent on a SERP nor the time spent on Web sites showed a significant effect on task success. As task and individual process characteristics jointly affect task success, the analyses were extended by exploring possible interactions between these characteristics. A significant negative interaction was found for the task characteristic number of information sources and the individual process characteristic number of visited Web sites ($b=-.17, p<.05$), implying that when completing tasks with a high number of information sources, the number of visited Web sites is less crucial for task success than when completing tasks with a low number of information sources.

As expected, having to compare a high number of information sources reduces the probability of task success, whereas congruency of credibility features increases the probability of task success. Both effects were discussed to be related to the amount of working memory capacity (Baddeley, 2012) that is needed to solve a task. The missing effect for the attractiveness of non-target search results was assumed to be due to task design. While it was found that the visit of many different Web sites increased the probability of task success, more did not always seem to be better. Concerning the missing effects for time, it was hypothesized that some individuals might spontaneously and unconsciously apply the right evaluation criteria leading to task success while others might need more time to consciously reflect on the evaluation process and credibility criteria. The interaction between the number of information sources and the number of visited web sites indicated that when evaluating a small number of information sources individuals benefit from using a highly elaborated information processing strategy, whereas with more of information sources, this strategy becomes less helpful because it may come at cost of mental effort (Rieh, Kim, & Markey, 2012).

8. General Discussion and Outlook

With the increasing and enormous importance of the Web in our everyday lives ICT skills have become indispensable for us in order to function and succeed in workplace settings, at school and at home. From basic skills like accessing information to cognitively more complex skills such as evaluating its quality, to successfully integrate ICT into our lives we need a set of different competencies. However, although there are many rationales to support the importance of ICT skills, their level still varies substantially between individuals, across countries and within communities (International ICT Literacy Panel, 2002). In order to comprehensively understand ICT literacy, develop reliable measures of ICT skills, enable us to generate effective training programs and lessen the digital divide, further research in this area is needed. Therefore, the purpose of this dissertation was to investigate two main ICT constructs, namely basic computer skills and the ability to evaluate the credibility of online information, from an individual differences perspective and introduce new interactive, computer-based measures for both. More precisely, the present dissertation aimed at further clarifying the theoretical and empirical specificity of these two ICT competencies by exploring their relation to selected cognitive covariates and investigating underlying processes of the evaluation of online-information applying the newly developed measurement tools.

In the following, the main findings of the three studies conducted are summarized and reflected on. Furthermore, methodological considerations and limitations of the present dissertation are outlined and an overview of this work's implications and possible directions for future research is given.

8.1 Summary and Discussion of Findings

As outlined in chapters 3 and 4 of this dissertation, the need of reliable, computer-based measures for basic computer skills and the ability to evaluate the credibility of online information was given in several aspects. In order to prepare students and adults for a successful life in a global economy increasingly dependent on ICT, there is a strong need of data helping us to understand not only how people successfully make use of the different ICT literacy skills but also recognize their structure and complexity (International ICT Literacy Panel, 2002). Accordingly, central finding of the first paper was that the newly developed BCS scale constitutes a reliable measurement tool. BCS seem to represent generic skills with BCS speed and BCS ability being specific constituents of BCS. In terms of the relation with related constructs, BCS showed strong associations with practical computer knowledge, which is in line with models like the IPS-I (Brand-Gruwel et al., 2009) model of problem solving, where basic computer skills are regarded as a requirement for the solution if higher-order tasks. It was also positively related with self-reported computer-skills. Basic reading skills such as word recognition were found to be associated with BCS performance concurrent with the fact that decoding of small text segments was required for the tasks of the BCS-scale. While the IPS-model assumes reading skills to be a basic requirement for problem solving on the internet, the results suggest that they are also a prerequisite for solving basic ICT tasks. Furthermore, BCS speed and ability were shown to predict electronic reading.

The second paper was concerned with the investigation of the theoretical background of the ability to evaluate the credibility of online information and the development of a correspondent assessment tool. It was shown that the ability to evaluate the credibility of online information, requiring the evaluation of structural and message features (ESMF), consists of two evaluation skills, namely ESF requiring the evaluation of structural features only, and EMF requiring the evaluation of message features only. In terms of the relation with selected cognitive covariates, it could be shown empirically that word recognition and reasoning ability were strongly related to ESF showing significant direct effects on ESF and significant indirect effects on ESMF via EFS. Thus, word recognition and reasoning ability can be assumed to be prerequisite skills for credibility evaluations of Web information. The results are in line with findings of the IPS-I model claiming that reading skills are conditional skills during the search for information on the internet. Furthermore, the positive effect of reasoning ability confirms earlier research by Coiro and Dobler (2007), who have shown that the ability to critically evaluate information involves processes of problem-solving, which are in turn predicted by reasoning ability (Funke & Frensch, 2007). In contrast to the expectations, basic computer skills were not revealed to contribute uniquely to ESMF and its constituents ESF and EMF, leading to the conclusion that trainings on how to evaluate online information do not necessarily have to incorporate training in computer skills because especially younger students may already possess sufficient computer experience (Aslanidou & Menexes, 2008).

The third paper explored terms of successful information evaluation on the Web and examined characteristics of online search results (task characteristics) and ways of interacting with them in this regard. As expected, concerning task characteristics, it was shown that having to compare a high number of information sources hinders a successful information evaluation, while congruency between credibility features in links on SERPs and corresponding Web sites facilitates the evaluation process. The negative effect of the number of information sources was discussed to be due to web users applying different information processing strategies as a function of external factors such as the number of decision alternatives (Wirth et al., 2007). In the case of a high number of information sources individuals were assumed to rely on heuristic processing rules like the “top link” heuristic (Pan et al., 2007), which might result in insufficient information evaluation and false conclusions. The positive effect of congruent credibility features was discussed in terms of working memory capacity. While the evaluation of conflicting information (incongruent credibility features) requires high working memory capacity, the processing of congruent credibility features was assumed to requires less working memory capacity for processes of information integration, thus providing more capacity for deeper evaluation processes. The attractiveness of non-targets did not show a significant effect on the successful evaluation of information, which was interpreted as result of the task design. Although the search results that had to be compared were very similar, the small number of search results still allowed a quite thorough and deep evaluation assumed to result in task success. Supporting this assumption, additional statistical analyses in fact revealed that users spent significantly more time on tasks with high attractiveness of non-targets than on tasks with low attractiveness of non-targets, presumably applying deep and systematic evaluation strategies. Concerning individual process characteristics it was shown that the strategy of visiting a high number of different Web sites was beneficial for identifying the most credible search result but became less efficient the more information sources needed to be compared. Contrary to the assumptions in this dissertation, the individual time invested in evaluating the credibility of search results on SERPs and Web sites did not have an effect on successfully identifying the most credible information piece. Due to quick and unconscious evaluation processes it was hypothesized that some individuals spontaneously and unconsciously apply the right evaluation criteria while others need more time to consciously reflect on the evaluation process.

To summarize, the three papers present reliable, interactive, computer-based measurement tools for basic computer skills and the ability to evaluate the credibility of online information. The exploration of cognitive covariates showed that word recognition can be considered a prerequisite skill for both ICT skills, while the ability to evaluate the credibility of online information was also shown to be associated with reasoning ability. Yet, although basic computer skills were significantly correlated with the ability to evaluate the credibility of online information (ESMF) and its constituents (ESF and EMF), they were not revealed to be a predictor of it as was assumed. A successful information evaluation on the Web was revealed to be fostered by visiting a high number of different information sources and the evaluation process was facilitated if credibility features in links on SERPs and corresponding web sites were congruent, while it was hindered if a high number of information sources needed to be compared. Finally, it should be acknowledged that the research presented in the three papers comes with certain limitations, which will be outlined in the following chapter.

8.2 Methodological Considerations and Generalizability of Findings

Several limitations need to be acknowledged when interpreting the main findings of this dissertation. The following section addresses methodological considerations with regard to the development of the measurement tools and the implementation of the studies. In addition it is reflected on the generalizability of the results.

Although the BCS scale presented in the first paper proved to be a reliable measurement tool, some methodological considerations need to be regarded. First, the BCS scale was not balanced with respect to simulated environments and information skills. Second, it showed a very restricted range of task difficulties. To obtain a more systematic construct representation and a wider range of task difficulties in order to also discriminate test takers with high levels of BCS ability, this was modified in the second paper by using a revised BCS scale incorporating additional tasks. With regard to the TEO test measuring the ability to evaluate the credibility of online information, a main limitation is the fact, that the three TEO subscales showed relatively low reliability coefficients (Cronbach's α). This might be due to the small number of tasks that each subscale contained (only 8 tasks per subscale). In addition effects might also be stronger with a larger amount of tasks. Therefore, the TEO subscales should be supplemented by additional tasks. Furthermore, although topics of the TEO tasks were selected to minimize the influence of background knowledge, it was not assessed and controlled for, which is strongly recommended for future research.

To draw conclusions from the presented research it is necessary to reflect the conditions under which the studies have been conducted. First of all, the three studies each used quite homogeneous samples with students either in grade 9 or grade 12. However, correlations and regression effects might be stronger with the use of more heterogeneous samples. Furthermore, the homogeneity of the samples limits the generalizability of the results, which are restricted to the students in grade 9 and 12 of the German school system.

Second, the studies come with several constraints: For the BCS scale participants solve tasks within a simulated computer environment that is not exhaustive with respect to the different possible approaches in solving the task that might be available with different existing computer programs. Concerning the TEO applied in paper 2 and 3 participants were supposed to work with an artificially designed SERP and a finite set of search results comprising either only 3 or 5 (subscales ESMF and EMF) or 6 or 10 (subscale ESF) search results, respectively. Furthermore, in the recent past the structure of SERPs has undergone several changes incorporating technological options that were not implied in the conducted studies. For example, images and videos are increasingly presented on SERPs in addition to text or auto completion lists giving several options for possible search terms already provide the Web user with feedback while entering single letters and words into the search field. In addition, TEO tasks only represent a small section of the Web search process according to Gerjets at al. (2011) not incorporating the need to recognize an information problem and define an adequate search term. These study conditions do not only limit the generalizability of the results to a broader range of users (e.g., older Web users or university students) but also to other contexts such as more natural search situations with real information needs, the possibility to enter and change own search terms and the wealth of search results returned by the search engine in a real Web search.

8.3 Implications and Future Research

The newly developed measurement tools BCS and TEO provide researchers with covariates for disentangling different sources of variance in performance of complex ICT tasks. From a practical point of view, they help to identify and measure what is done to lessen the digital divide and advance the development of adequate training programs and interventions that help school kids, university students, and working adults to improve their basic computer skills and inform them on how to successfully evaluate the credibility of information on the Web (e.g., Gerjets & Hellenthal-Schorr, 2008; Walraven, Brand-Gruwel, & Boshuizen, 2010; Wiley et al., 2009). The present work suggests that besides teaching individuals about information credibility criteria, trainings in ICT skills might also benefit from including tasks improving basic cognitive skills such as word recognition and basic reading skills. Furthermore, findings indicate effective evaluation strategies that can be incorporated in Web search trainings. For example a helpful evaluation strategy during Web search seems to be the visit of many different Web sites, although this strategy becomes less effective when it exceeds a certain number of information sources.

Future research in this area is needed to explore the relation of basic computer skills and the ability to evaluate the credibility of online information with the other ICT competencies defined by the International ICT literacy Panel (2002) for sake of a comprehensive understanding of ICT literacy. Furthermore, it should be interesting to explore the relation of both constructs with other cognitive covariates. As word recognition was found to have associations with both constructs it might be rewarding to investigate the relation of different higher-order facets of reading literacy with both. Research in the field of reading literacy has revealed that poor readers tend to focus on the decoding of individual words at the expense of deeper comprehension (Perfetti, 1985). Therefore, there might not be a difference between poor and fluent readers in BCS performance and the subscale ESF of the TEO. However, higher-order facets of reading literacy might have a strong effect on the subscales ESMF and EMF because they involve reading in-depth. Moreover, while Kornmann et al. (2016) have shown that spatial working memory is crucial for an effective navigation in multiperspective hypermedia environments (MHEs), the role of working memory capacity should also be explored in relation to a successful evaluation of online information. With regard to the TEO future research should incorporate the methodology of eye-tracking and thinking-aloud protocols to gain deeper insights of the processes carried out during Web search (Cutrell & Guan, 2007; Kammerer & Gerjets, 2012; Van Gog, Paas, & van Merriënboer, 2005) that go beyond overt actions including information about which credibility features individuals knowingly make use of when judging the credibility of online information. The application of eye-tracking could also help to track navigation paths during Web search and identify ideal navigation paths enabling us to recommend general guidelines for successful information evaluation on the Web.

Despite the limitations outlined, the research presented does not only introduce new, reliable, interactive and computer-based measures for basic computer skills and the evaluation of the credibility of online information, but also provides novel insights into their latent structure, their relation with each other and further cognitive covariates, as well as terms of successful information evaluation on the Web with regard to task and individual process characteristics. In conclusion, the present dissertation addresses a major topic of the 21st century and contributes to a deeper and differentiated understanding of ICT literacy in general hopefully bringing forward future research fostering a conscious and competent handling of the possibilities that ICT establishes.

References

Alexander, J.E., & Tate, M.A. (1999). *Web wisdom: How to evaluate and create information quality on the Web*. Hillsdale, NJ: Erlbaum.

Anderson, J.R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*, 369–406.

Anderson, J.R., Matessa, M., & Lebiere, C. (1997). ACT-R: A theory of higher level cognition and its relation to visual attention. *Human-Computer Interaction, 12*, 439–462.

Antos, Gerd (2017): Fake News. Warum wir auf sie reinfallen. Oder: „Ich mache euch die Welt, so wie sie mir gefällt“. *Der Sprachdienst 1/17*. S. 1-20.

Aslanidou, S., & Menexes, G. (2008). Youth and the internet: Uses and practices in the home. *Computers and Education, 51*, 1375–1391.

Baddeley, A. D. (2012). Working memory: theories, models, and controversies. Annual *Review of Psychology, 63*, 1-29.

Balota, D.A., Cortese, M.J., Sergent-Marshall, S.D., Spieler, D.G., & Yap, M.J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133*, 283-316.

Bates, D., Maechler, Martin, & Bolker, B. (2011). lme4: Linear mixed-effects models using S4 classes. R package version 1.1-13. http://CRAN.R-project.org/package=lme4

Bollen, K.A. (1989). *Structural equations with latent variables*. New York: Wiley.

Brand-Gruwel, S., Wopereis, I., & Walraven, A. (2009). A descriptive model of Information Problem Solving while using Internet. *Computers & Education, 53*, 1207-1217.

Brandt, D.S. (1996). Evaluating information on the Internet. *Computers in Libraries, 16*, 44-46.

Bromme, R. (2005). Thinking and knowing about knowledge: a plea for and critical remarks on psychological research programs on epistemological beliefs. In M. Hoffmann, J. Lenhard, & F. Seeger (Eds.), *Activity and sign - Grounding mathematics education* (pp. 191-201). New York: Springer.

Carroll, J.B. (1993). *Human cognitive abilities*. New York, NY: Cambridge University Press.

Castek, J., Zawilinski, L., McVerry, G., O'Byrne, I., & Leu, D. (2011). The new literacies of online reading comprehension: New opportunities and challenges for students with learning difficulties. In: C. Wyatt-Smith, J. Elkins, & S. Gunn (Eds.) *Multiple Perspectives on Difficulties in Learning Literacy and Numeracy* (pp. 91-110). New York, NY: Springer.

Chaiken, S., Liberman, A., & Eagly, A. H. (1989). Heuristic and systematic processing within and beyond the persuasion context. In J. Uleman & J. Bargh (Eds.), *Unintended Thought* (pp. 212–252). New York: Guilford Press.

Chen, S., & Chaiken, S. (1999). The heuristic-systematic model in its broader context. In S. Chaiken & Y. Trope (Eds.), *Dual-Process Theories in Social Psychology* (pp. 73–96). New York: Guilford Press.

Coiro, J., & Dobler, E. (2007). Exploring the online reading comprehension strategies used by sixth-grade skilled readers to search for and locate information on the internet. *Reading Research Quarterly, 42*, 214-257.

Coiro, J., Knobel, M., Lankshear, C., & Leu, D. (2008). Central issues in new literacies and new literacies research. IN: J. Coiro, M. Knobel, C. Lankshear and D. Leu (Eds.) *Handbook of research on new literacies* (pp. 1-31). New York: Lawrence Erlbaum Associates.

Cutrell, E., & Guan, Z. (2007). What are you looking for? An eye-tracking study of information usage in Web search. *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, 407-416.

Evans, J. S. (2008). Dual-processing accounts of reasoning, judgment and social cognition. *Annual Review of Psychology, 59*, 255-278.

Fogg, B.J. (2002). *Prominence-Interpretation Theory: Explaining How People Assess Credibility*. A Research Report by the Stanford Persuasive Technology Lab, Stanford University. Available at http://credibility.stanford.edu/pit.html or http://www.webcredibility.org/pit.html.

Fritch, J.W., & Cromwell, R.L. (2001). Evaluating Internet resources: Identity, affiliation, and cognitive authority in a networked world. *Journal of the American Society for Information Science and Technology, 52* (6), 499-507.

Funke, J., & Frensch, P.A. (2007). Complex problem solving: The European perspective - 10 years after. In: David H. Jonassen (Ed.) *Learning to solve complex scientific problems* (pp. 25-47). New York: Lawrence Erlbaum Associates. Available at: http://works.bepress.com/joachim_funke/18

Gerjets, P., & Hellenthal-Schorr, T. (2008). Competent information search in the World Wide Web: development and evaluation of a Web training for pupils. *Computers in Human Behavior, 24*, 693-715.

Gerjets, P., Kammerer, Y., & Werner, B. (2011). Measuring spontaneous and instructed evaluation processes during Web search: Integrating concurrent thinking-aloud protocols and eye-tracking data. *Learning and Instruction, 21*, 220-231.

Hargittai, E., Fullerton, L., Menchen-Trevino, E., & Thomas, K. Y. (2010). Trust Online: Young Adults' Evaluation of Web Content. *International Journal of Communication, 4*, 468-494.

Hilligoss, B., & Rieh, S. (2008). Developing a unifying framework of credibility assessment: Construct, heuristics, and interaction in context. *Information Processing and Management, 44*, 1467-1484.

Hong, T. (2006). The influence of structural and message features on website credibility. *Journal of the American Society for Information Science and Technology, 57* (1), 114-127.

Ilomäki, L.,& Rantanen, P. (2007). Intensive use of ICT in school: Developing differences in students' ICT expertise. *Computers and Education, 46*, 119–136.

International ICT Literacy Panel (2002). *Digital transformation: A framework for ICT literacy (A report of the International ICT Literacy Panel)*. Princeton, NJ: Educational Testing Service. Retrieved from http://www.ets.org/Media/Research/pdf/ictreport.pdf

Kammerer, Y., & Gerjets, P. (2012). Effects of search interface and Internet-specific epistemic beliefs on source evaluations during Web search for medical information: an eye-tracking study. *Behaviour & Information Technology, 31* (1), 83-97.

Kammerer, Y., & Gerjets, P. (2014). The Role of Search Result Position and Source Trustworthiness in the Selection of Web Search Results When Using a List or a Grid Interface. *Computer-Interaction, 30*, 177-191.

Kapoun, J. (1998). *Teaching undergrads WEB evaluation.* A guide for library instruction. Available from: http://www.ualberta.ca/~dmiall/Brazil/Kapoun.html [Accessed 17 October 2012].

Keßel, Y., Goldhammer, F., & Kröhne, U. (under review). Evaluating the Credibility of Online Information: The influence of basic computer skills, word recognition and reasoning.

Kornmann, J., Kammerer, Y., Anjewierden, A., Zettler, I., Trautwein, U., & Gerjets, P. (2016). *Computers in Human Behavior, 55*, 145-158.

Kuhlemeier, H., & Hemker, B. (2007). The impact of computer use at home on students' internet skills. *Computers and Education, 49*, 460–480.

Leu, D.J., Kinzer, C.K., Coiro, J.L., & Cammack, D.W. (2004). Toward a theory of new literacies emerging from the internet and other information and communication technologies. In N.J. Unrau & R.B. Ruddell (Eds.), *Theoretical models and processes of reading* (5th ed., pp.1570–1613). Newark, NJ: International Reading Association.

Lewandowski, D. (2011). The influence of commercial intent of search results on their perceived relevance. In S. Sadagopan & K. Ramamritham (Eds.), *Proceedings of the 2011 iConference* (iConference '11; pp. 452–458). NewYork, NY: ACM Press.

Lewandowski, D. (2013). Credibility in Web search engines. In S. Apostel & M. Folk (Eds.), *Online credibility and digital ethos: Evaluating computer-mediated communication* (pp. 131–146). Hershey, PA: IGI Global

Lorenzen, M. (2001). The land of confusion? High school students and their use of the World Wide Web for research. *Researching Strategies, 18*, 151-163.

Lucassen, T., & Schraagen, J.M. (2011). Factual accuracy and trust in information: The role of expertise. *Journal of the American Society for Information Science and Technology, 62* (7), 1232-1242.

MaKinster, J. G., Beghetto, R. A., & Plucker, J. A. (2002). Why can't I find Newton's third law? Case studies of students' use of the web as a science resource. *Journal of Science Education and Technology, 11*, 155–172.

Markauskaite, L. (2007). Exploring the structure of trainee teachers' ICT literacy: The main components of and relationships between, general cognitive and technical capabilities. *Educational Technology Research and Development, 55*, 547–572.

Mason, L., Junyent, A. A., & Tornatora, M. C. (2014). Epistemic evaluation and comprehension of web-source information on controversial science-related topics: Effects of a short-term instructional intervention. *Computers & Education, 76*, 143–157.

Mayer, R.E. (2003). What causes individual differences in cognitive performance? In R. J. Sternberg & E.L. Grigorenko (Eds.), *The psychology of abilities, competencies, and expertise* (pp. 263–74). New York, NY: Cambridge University Press.

Meola, M. (2004). Chucking the checklist: A contextual approach to teaching undergraduates Web-site evaluation. *Libraries and the Academy, 4* (3), 331-344.

Metzger, M. (2007). Making Sense of Credibility on the Web: Models for Evaluating Online Information and Recommendations for Future Research. *Journal of the American Society for Information Science and Technology, 58*(13), 2078-2091.

Metzger, M.J., Flanagin, A.J., & Zwarun, L. (2003). College student Web use, perceptions of information credibility, and verification behavior. *Computers & Education, 41*, 271-290.

Muthén, B.O. (1998-2004). *Mplus Technical Appendices*. Los Angeles, CA: Muthén & Muthén.

Muthén, L. K. & Muthén, B. O. (1998-2011). *Mplus User's Guide*. Sixth Edition. Los Angeles, CA: Muthén & Muthén.

OECD (2011). *PISA 2009 Results: Students on Line: Digital Technologies and Performance* (Volume VI). Available from: http://dx.doi.org/10.1787/9789264112995-en [Accessed 17 October 2012].

Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G., & Granka, L. (2007). In Google we trust: Users' decisions on rank, position and relevance. *Journal of Computer-Mediated Communication, 12*, 801-823.

Parshall, C.G., Spray, J.A., Kalohn, J.C., & Davey, T. (2002). *Practical considerations in computer-based testing*. New York, NY: Springer.

Perfetti, C.A. (1985). *Reading ability*. New York: Oxford University Press.

Petty, R. E., & Cacioppo, J. T. (1986). The Elaboration Likelihood Model of Persuasion. In: *Advances in experimental social psychology* (Ed. L. Berkowitz), 19, pp. 123 – 205. New York: Academic Press.

Petty, R. E., & Cacioppo, J. T. (1990). *Communication and Persuasion*. New York: Springer.

R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Richter, T., Naumann, J., & Horz, H. (2010). Eine revidierte Fassung des Inventars zur Computerbildung (INCOBI-R) [A revised version of the Computer Literacy Inventory]. Zeitschrift für Pädagogische Psychologie/ German Journal of *Educational Psychology, 24*, 23–37.

Richter, T., Naumann, J., Isberner, M.-J., & Kutzner, Y. (2011). Diagnostik von Lesefähigkeiten bei Grundschulkindern: Eine prozessorientierte Alternative zu produktorientierten Tests [Assessment of reading skills in primary school children: A process-oriented alternative to product-oriented tests]. *Diskurs Kindheits- und Jugendforschung, 6*, 479-486.

Rieh, S.Y. (2002). Judgment of information quality and cognitive authority in the Web. *Journal of the American Society for Information Science and Technology, 53*(2), 145-161.

Rieh, S. Y., Kim, Y. M., & Markey, K. (2012). Amount of invested mental effort (AIME) in online searching. *Information Processing and Management, 48* (6), 1136–1150.

Rizopoulos, D. (2006). *ltm: An R Package for Latent Variable Modeling and Item Response Theory Analyses.* [online] Journal of Statistical Software, 17(5) 1-25. Available from: http://www.jstatsoft.org/v17/i05/ [Accessed 17 October 2012].

Rouet, J.-F., Bétrancourt, M., Britt, A. M., Bromme, R., Graesser, A. R., Kulikowich, J. M., Leu, D. J., Ueno, N., & van Oostendorp, H. (2009). PIAAC problem solving in technology-rich environments: a conceptual framework. (OECD education working paper No. 36). Retrieved December 14, 2009 from. http://www.oecd.org/edu/workingpapers.

Russell, M., Goldberg, A., & O'Connor, K. (2003). Computer based testing and validity: A look back into the future. *Assessment in Education: Principles, Policy and Practice, 10*, 279–293.

Salméron, L., Kammerer, Y., & García-Carríon, P. (2013). Searching the Web for conflicting topics: Page and user factors. *Computers in Human Behavior, 29*, 2161-2171.

Salmerón, L., Naumann, J., García, V., & Farjado, I. (2017). Scanning and deep processing of information in hypertext: an eye tracking and cued retrospective think-aloud study. *Journal of Computer Assisted Learning, 33*, 222–233.

Sandene, B., Bennett, R.E., Braswell, J., & Oranje, A. (2005). Online assessment in mathematics. In B. Sandene, N. Horkay, R.E. Bennett, N. Allen, J. Braswell, B. Kaplan, & A. Oranje (Eds.), *Online assessment in mathematics and writing: Reports from the NAEP Technology-based Assessment Project (NCES 2005–457)* (pp.1–68). Washington, DC: US Department of Education, National Centre for Education Statistics.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Test of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research - Online, 8*, 23-74.

Schipolowski, S., Schroeders, U., & Wilhelm, O. (2008). BEFKI - Berlin Test of Fluid and Crystallized Intelligence. Poster presented at the XXIX International Congress of Psychology, Berlin, Germany.

Scholz-Crane, A. (1998). Evaluating the future: A preliminary study of the process of how undergraduate students evaluate Web sources. *Reference Services Review, 26* (3/4), 53-60.

Schwarz, J., & Morris, M. R. (2011). Augmenting Web Pages and Search Results to Support Credibility Assessment. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, May 07-12, 2011, Vancouver, BC, Canada [doi>10.1145/1978942.1979127]

Smith, A.G. (1997). *Testing the surf: Criteria for evaluating Internet information resources.* [online] Public-Access Computer Systems Review, 8. Available from http://journals.tdl.org/pacsr/article/viewFile/6016/5645 [Accessed 17 October 2012].

Tu, Y., Shih, M., & Tsai, C. (2008). Eighth graders' web searching strategies and outcomes: The role of task types, web experiences and epistemological beliefs. *Computers and Education, 51*, 1142–1153.

Van Gog, T., Paas, F., & van Merriënboer, J. J. G. (2005). Uncovering expertise-related differences in troubleshooting performance: combining eye movement and concurrent verbal protocol data. *Applied Cognitive Psychology, 19*, 205-221.

Wallace, R. M., Kupperman, J., Krajcik, J., & Soloway, E. (2000). Science on the web: Students online in a sixth-grade classroom. *The Journal of the Learning Sciences, 9*, 75–104.

Walraven, A., Brand-Gruwel, S., & Boshuizen, H. P. A. (2009). How students evaluate information and sources when searching the World Wide Web for information. *Computers and Education, 52*, 234–246.

Walraven, A., Brand-Gruwel, S., & Boshuizen, H. P. A. (2010). Fostering transfer of websearchers' evaluation skills: A field test of two transfer theories. *Computers in Human Behavior, 26*, 716–728.

Wathen, C.N., & Burkell, J. (2002). Believe it or not: Factors influencing credibility on the Web. *Journal of the American Society for Information Science and Technology, 53* (2), 134-144.

Wecker, C., Kohnle, C., & Fischer, F. (2007). Computer literacy and inquiry learning: When geeks learn less. *Journal of Computer-Assisted Learning, 23*, 133–144.

Weinert, S., Artelt, C., Prenzel, M., Senkbeil, M., Ehmke, T., & Carstensen, C.H. (2011). Development of competencies across the life span. *Zeitschrift für Erziehungswissenschaft, 14*, 67-86.

Wiley, J., Goldman, S., Graesser, A., Sanchez. C., Ash, I., & Hemmerich, J. (2009). Source evaluation, comprehension, and learning in internet science inquiry tasks. *American Educational Research Journal. 46*, 1060-1106.

Wirth, W., Böcking, T., Karnowski, V., & von Pape, T. (2007). Heuristic and systematic use of search engines. *Journal of Computer-Mediated Communication, 12*, 778-800.

Paper 1

Publication Note

Goldhammer, F., Naumann, J., & Keßel, Y. (2013). Assessing Individual Differences in Basic Computer Skills. *European Journal of Psychological Assessment, 29* (4), 263-275.

Assessing Individual Differences in Basic Computer Skills
Psychometric Characteristics of an Interactive Performance Measure

A definition of basic computer skills (BCS) is proposed and the psychometric properties of a newly developed BCS scale are investigated. BCS is defined as the ability and speed of performing basic actions in graphical user interfaces of computers to access, collect, and provide information. BCS is thus considered a basic component skill of the much broader construct of ICT literacy. Data from the German PISA 2009 field trial was used to determine the factor structure of the BCS scale as well as convergent and discriminant validity. The latent factor structure underlying the BCS scale was investigated by testing confirmatory factor analysis (CFA) models for response times and responses. CFA results suggest that there is one dimension of BCS speed and BCS ability, respectively. With respect to convergent validity, practical computer knowledge and skill in digital reading had strong associations with BCS speed and ability. With respect to discriminant validity, only moderate associations were found with lower level reading skills and self-reported computer skills. Differences between BCS speed and ability and further developments of the BCS scale are discussed.

1. Introduction

In this paper, we describe the development and validation of a scale assessing basic computer skills. Given the increasing importance of information and communication technology (ICT) in peoples' everyday lives, computer-related skills have become a significant research topic from the perspective of individual differences and assessment (e.g., Kim, Jung, & Lee, 2008; Lennon, Kirsch, Von Davier, Wagner, & Yamamoto, 2003; Poynton, 2005; Russell, Goldberg, & O'Connor, 2003). Many conceptualizations of ICT literacy are available which differ in the extent to which they focus on technological, informational, and motivational aspects. Tsai (2002) defines computer literacy in a broad sense as "the basic knowledge, skills, and attitudes needed by all citizens to be able to deal with computer technology in their daily life" (p. 69). In the present study, ICT literacy is conceptualized as a competence (see Klieme, Hartig, & Rauch, 2008), i.e., as a context-specific cognitive disposition including knowledge, skills, and routines that are acquired by learning, and that are required to deal successfully with informational tasks in ICT contexts. In this sense, the International ICT Literacy Panel (2002) presents a definition of ICT literacy which does not take attitudes into account, but focuses on cognitive and technological aspects. The Panel suggests five information-related competencies constituting ICT literacy:

> ICT literacy is using digital technology, communications tools, and/or networks to access, manage, integrate, evaluate, and create information in order to function in a knowledge society (p. 16).

This article presents a newly developed scale for the assessment of basic ICT skills utilizing interactive simulations. Basic ICT skills are not to be confounded with ICT literacy, but can be seen as the lower-level part of ICT literacy. Models of computer-related information processing like the IPS-I model (Information Problem Solving on the Internet; Brand-Gruwel, Wopereis, & Walraven, 2009) distinguish between three layers of component skills and processes. The lowest level is made up of basic skills, such as basic reading and computer skills. These skills are required for cognitive actions of searching, processing, and integrating information to be performed on the second layer. Finally, on the third layer, information processing is governed by metacognitive regulation processes, such as monitoring and evaluating. It is important to note that the scale described in this work addresses only the first layer – and here only basic computer-related skills.

Consider, for example, a student required to find information for an essay on genetically engineered foods. This student might engage in an internet search to this end, e.g., by using a search engine, requiring him or her to access, read, evaluate, and integrate multiple digital documents to solve the information problem. Carrying out such a complex ICT-based information search requires a number of basic component skills. Obviously, it is necessary to be able to read, but also basic knowledge and related perceptual and motor skills needed to use the computer interface are required. For instance, the box in the search engine has to be identified, knowledge about the correct keys or buttons to send the search request must be available, knowledge about how to operate hyperlinks and a browser interface must be known, etc. In addition, using the technical interface should happen as quickly and automatically as possible, in order to free up cognitive capacity on higher layers, e.g., evaluating source credibility when integrating information from various sources. Throughout this article, when we refer to basic computer skills, we mean lower-level skills only that are highly automatic and usually represented as procedural rather than declarative knowledge (see Anderson, 1982).

We relate our conceptualization of basic computer skills (BCS) to other existing frameworks and argue for a performance-based measure of basic ICT skills which takes into account not only accuracy, but also the response times in performing basic ICT tasks. Then we explicate and test our hypotheses about the psychometric properties of the BCS scale and its validity, i.e., we focus on the dimensionality of BCS speed and BCS ability as well as their covariance structure. Here, we follow the approach of confirmatory measurement models and conceive speed and ability as latent variables measured by actual response times and response accuracy, respectively (see, e.g., van der Linden, 2009). Finally, we investigate how BCS are related to other cognitive and demographic variables, in order to determine the BCS scale's convergent and discriminant validity.

1.1 Definition of Basic Computer Skills

We propose conceptualizing the construct of basic computer skills (BCS) as *the fundamental ability and speed of performing basic actions in graphical user interfaces of computers to access, collect, and provide information.*

The definition refers to fundamental skills that can be considered core technical skills enabling one to perform simple actions being common to many software applications, e.g., handling the file-management function (see Markauskaite, 2007). Such actions can be completed in few steps and are usually encountered regardless of the current reason for using a particular software, e.g., operating the copy-and-paste functionality of a word processor may be used when writing a letter but also when creating a timetable.

Basically, *ability* and *speed* represent major aspects of human performance in cognitive tasks (for a review, see Carroll, 1993). From a measurement perspective, ability and speed are hypothesized person parameters to explain individual differences in response behavior above and beyond the influence of task characteristics, i.e., speed differences account for response time differences and ability differences explain differences in accuracy (see van der Linden, 2009; for an example, see Goldhammer & Klein Entink, 2011). Considering both speed and ability, which together indicate efficiency, seems most appropriate when focusing on basic skills reflected by relatively easy tasks (see the approach by Sandene, Bennett, Braswell, & Oranje, 2005, measuring both input speed and input accuracy as components of computer skills). The definition centers on *information*, i.e., ICT is considered as a tool applicable to dealing with information for solving tasks of daily life. The term *computer* is used in a general sense and covers the operation of desktop computers, which may be connected to networks, but also communication devices. The human computer interface is assumed to be *graphical* given the importance of visual perception in human information processing (e.g., Anderson, Matessa, & Lebiere, 1997). To represent programs and commands graphically, the user interface includes windows, icons, and menus that can be manipulated by a mouse and keyboard. *Accessing* information involves basic technical skills for making information available for acquisition and further processing, e.g., using links in a web environment, navigation buttons of a web browser, menus, and search functions. *Collecting* information is related to basic ways of gathering information to maintain accessibility, e.g., file management, creating bookmarks and connections, typing, as well as copying and pasting. Finally, *providing* information means information is made accessible for others, using basic technical functions for sharing information, e.g., by sending or forwarding e-mail messages in an e-mail client.

The relevance of the construct of BCS and a valid measure of BCS is given in several respects. First, BCS itself is a highly relevant competence construct, in that individuals lacking sufficient basic ICT skills will clearly be left behind both in educational as well as in professional contexts (see Leu, Kinzer, Coiro, & Cammack, 2004). Related research requires measures of ICT literacy to assess it cross-sectionally or longitudinally, e.g., to study the effect of instructional interventions aiming at the improvement of ICT literacy or to investigate the development of ICT literacy across the lifespan (e.g., Weinert et al., 2011). Second, the BCS construct has become an important control variable because numerous assessments increasingly rely on computer-based assessment procedures. If the individual level of prior computer experience affects the test score above and beyond the cognitive disposition to be assessed, the validity of the test score is threatened (e.g., Parshall, Spray, Kalohn, & Davey, 2002; Russell et al., 2003; Sandene et al., 2005). Thus, when developing computer-based assessments of cognitive dispositions, it is methodologically crucial to minimize the influence of individual differences in computer-related skills or control by assessing them as covariate. Third, the computer itself has become an important means of instruction, and materials to be learned are delivered through the computer as hypertext, hypermedia, or simulations. In research on learning with the computer, ICT skills are generally an important covariate to be measured (e.g., Wecker, Kohnle, & Fischer, 2007).

1.2 Assessment of Basic Computer Skills

Self-ratings of ICT skills have often been used in their assessment (e.g., the ICT Self-Efficacy Scale; Markauskaite, 2007). Although self-ratings provide some insight into actual skills (e.g., Ballantine, Larres, & Oyelere, 2007), they also reveal some shortcomings. For example, students tend to overrate their actual skill in some ICT domains, such as adequately judging the reliability of information found on the internet (Metzger, Flanagin, & Zwarun, 2003). Furthermore, some evidence points to this effect being more pronounced for boys than for girls (Hakkarainen et al., 2000), possibly causing an overestimation of boys' advantages over girls in ICT-related skills.

In another approach to the measurement of ICT literacy, paper-and-pencil-based tests including multiple-choice questions were used. For example, in their Internet Skills for School test, Kuhlemeier and Hemker (2007) used screenshots depicting a task and asked students what best to do in that particular situation. A similar approach was taken by Potosky (2007) with the Internet Knowledge (iKnow) measure. Similarly, in their Computer Literacy Inventory (INCOBI), Richter, Naumann, and Horz (2010) used verbal scenarios of everyday computer tasks, together with four response options, only one of which was correct.

Although paper-and-pencil tests of computer knowledge are more objective – and are better than self-reports in the prediction of actual performance – they also have limitations. Most important, unlike computer simulations, paper-and-pencil tests cannot provide an authentic interactive task. This is a major drawback if computer-related skills need to be assessed. Moreover, paper-and-pencil tests usually do not provide information about the test takers' response times. However, in the case of basic and, therefore, easy computer skills tasks, individual differences might be even greater with respect to speed of task completion than to response accuracy. Following this reasoning, the present study aimed at the development of an interactive and objective performance measure to assess ability and speed in interacting with a simulated computer environment. Although this strategy seems to be most adequate for assessing ICT literacy conceptualized as an interactive competence, only a few performance-oriented measures are available, e.g., the Information and Communication Technology Literacy Test (iSkills) by ETS (2008; see also Katz & Macklin, 2007).

The goal of the present work was to develop a BCS scale for the national extension of the Program for International Student Assessment (PISA) 2009 study targeting German 15-year-old secondary school students. We designed tasks that simulate typical computer environments and require responses within these environments using mouse and/or keyboard.

1.3 Hypotheses

The first hypothesis addresses the psychometric properties of the newly developed BCS scale:

Hypothesis 1: Since BCS enable us to deal with task requirements that are common to many software applications and purposes, we assume unidimensionality in both BCS speed and BCS ability. The unidimensional model is contrasted with a plausible alternative, less restricted measurement model assuming environment-specific dimensions, i.e., ability and speed dimensions being specific to text editor, web-browser and e-mail client. The two constructs are expected to be positively related dimensions, i.e., more able participants tend to complete basic ICT tasks faster.

The following hypotheses refer to the convergent and discriminant validity of the BCS scale. The construct validity of the BCS scale is addressed by investigating how well covariates assumed to underlie BCS actually predict BCS (Hypothesis 2.1), by clarifying the relationship of the objective BCS scale to a subjective scale assessing computer skills (Hypothesis 2.2), by replicating gender associations with computer-related skills (Hypothesis 2.3), and by investigating the predictive validity of BCS as an underlying component of electronic reading ability (Hypothesis 2.4).

Hypothesis 2.1a: Domain knowledge is assumed to be a necessary condition for literate behavior in that competence acquisition is considered a learning process that includes gaining and applying knowledge (see Mayer, 2003). For BCS, we assume that particularly practical computer knowledge, i.e., the knowledge about how to solve everyday computer problems, facilitates the development of BCS. The relationship of BCS and computer knowledge, however, is probably bidirectional: While good knowledge helps one to develop good BCS, good BCS also helps in operating the computer without being under cognitive strain and thus helps in the acquisition of new knowledge of procedures (see, e.g., Keith & Frese, 2005; Keith, Richter, & Naumann, 2010).

As a consequence, we assume that practical computer knowledge has strong associations with BCS ability and BCS speed, respectively.

Hypothesis 2.1b: While it is largely undisputed that ICT literacy generally has reading skills as a basis (among others), this is less clear for BCS. However, BCS tasks also usually include detecting and reading simple verbal labels, e.g., buttons, menu items. Thus, higher level components of reading skill, such as integrating text contents with prior knowledge to form a mental model (see, e.g., Kintsch, 1998), are assumed to be distinct from BCS. However, lower level decoding skills, measured by tasks such as lexical decision, are required. We thus assume that ability and speed in word recognition are also associated with BCS performance.

Hypothesis 2.1c: When taking both practical computer knowledge and word recognition into account, we assume effect sizes found for word recognition to be smaller than those for practical computer knowledge, because basic reading abilities are a necessary, albeit not sufficient, condition for the successful completion of BCS tasks.

Hypothesis 2.2: BCS ability and speed are assumed to be positively related to self-reported computer skills (see, e.g., Ballantine et al., 2007). Given that the assessment of BCS and computer skills involve two different types of measures (i.e., objective test vs. subjective self-report) and do not measure exactly the same construct (i.e., computer skills refer to more complex ICT activities than BCS, and reflect not only skills, but also self-efficacy), the correlations between BCS ability and speed are assumed to be moderate, that is, at most about .30. This magnitude is usually found when self-report measures are compared to performance measures of the same ability construct (see Mabe & West, 1982).

Hypothesis 2.3: A number of recent studies indicated that male students have (slightly) better ICT skills than females (e.g., Ilomäki & Rantanen, 2007; Imhof, Vollmeyer, & Beierlein, 2007; Kuhlemeier & Hemker, 2007). As the development of ICT skills depends on learning and opportunities to learn, we do not assume that males are principally superior to females. However, males tend to use the computer more often and more intensively than females (e.g., Colley & Comber, 2003; Schumacher & Mohran-Martin, 2001; Vekiri & Chronaki, 2008), giving them more practice, which in turn might produce faster and more accurate responses especially in lower level tasks. Based on this assumption, a valid BCS measure should be able to reveal the (slight) advantage of males. Therefore, we assume that male participants show higher mean levels of BCS ability and speed than females.

Hypothesis 2.4: Finally, we assume BCS ability and speed to be predictors of electronic reading ability. Compared to traditional reading literacy, reading electronic texts presumably affords additional and specific processing requirements because printed and electronic texts differ in structure. Most important, electronic reading requires people to deal with the computer interface, so that BCS are needed, for instance, to navigate in a web-browser (see Leu et al., 2004; OECD, 2011).

2. Material and Methods

2.1 Participants

A total of 320 secondary school students were included in the study. They participated in the German field trial of the PISA 2009 study, the target population consisting of 15year-old students. Five participants with very high missing rates in BCS data (i.e., they completed less than one third of the 15 BCS tasks) were excluded from data analysis, resulting in a sample size of 315. The sample included 50.5% females and 42.5% males (7% unknown), aged 15.42 to 16.33 years (M = 15.87, SD = 0.28).

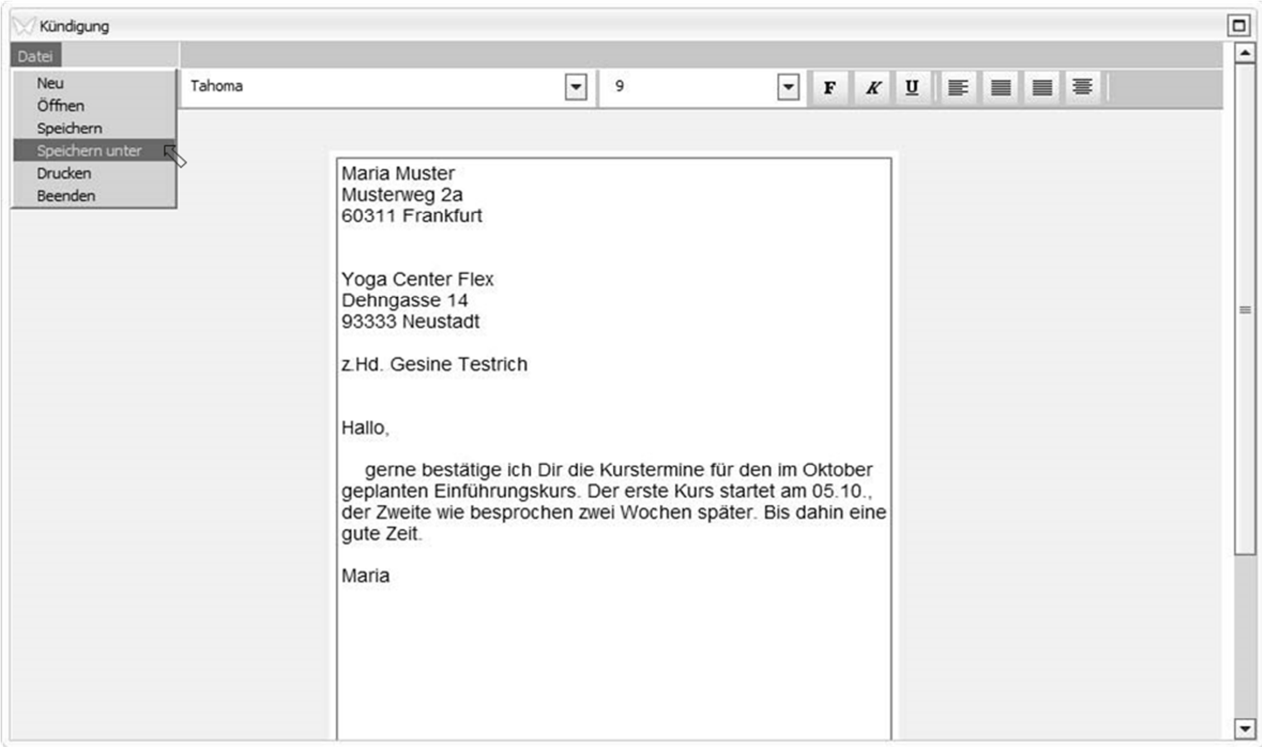
2.2 Measured Variables

2.2.1 Basic computer skills.

The Basic Computer Skills Test comprised 15 tasks (see Table 1), designed to cover the construct of BCS as defined above. Seven tasks basically require students to access information (1, 2, 5, 6, 9, 10, 11), six tasks require them to collect information (3, 4, 7, 8, 12, 14), and two tasks focus on providing information (13, 15). Each task starts with a brief instruction explaining the BCS task embedded in an informational context. Then, the stimulus including a simulated graphical user interface of a computer environment is presented (see examples in Figure 1). The tasks refer to either a web environment (nine tasks), a text editor environment (four tasks), or an e-mail client environment (two tasks) (see Table 1). The focus is on web environments, as browsing the internet can be considered a major computer-related activity in the targeted age group (e.g., OECD, 2005, 2011). The simulated environments were abstracted from real software and operating systems. However, they can be assumed to share general characteristics of interaction functions with real computer environments (e.g., clicking onto a menu opens a list of menu items, clicking onto blue underlined text in a web browser links to another page, etc.). All tasks are supposed to be solved by using the mouse, except for tasks 4 (Typing) and 10 (Finding a string), which require use of the keyboard for entering text. Nine tasks can be solved by a single interaction with the environment, while five tasks require two steps for completion, and one task requires three steps (see Table 1, column required interactions to solve the task).

Test takers are asked to complete the tasks as quickly and accurately as possible. They navigate between tasks by clicking a “next” button to the left of the stimulus area. This is practiced in a tutorial at the beginning of the test. We took care that the BCS tasks were delivered in a secure testing environment, allowing for the logging of user interactions, including time stamps to compute response times. For each task, we collected the individual log-transformed response time (RT) and the response (R).

A



B

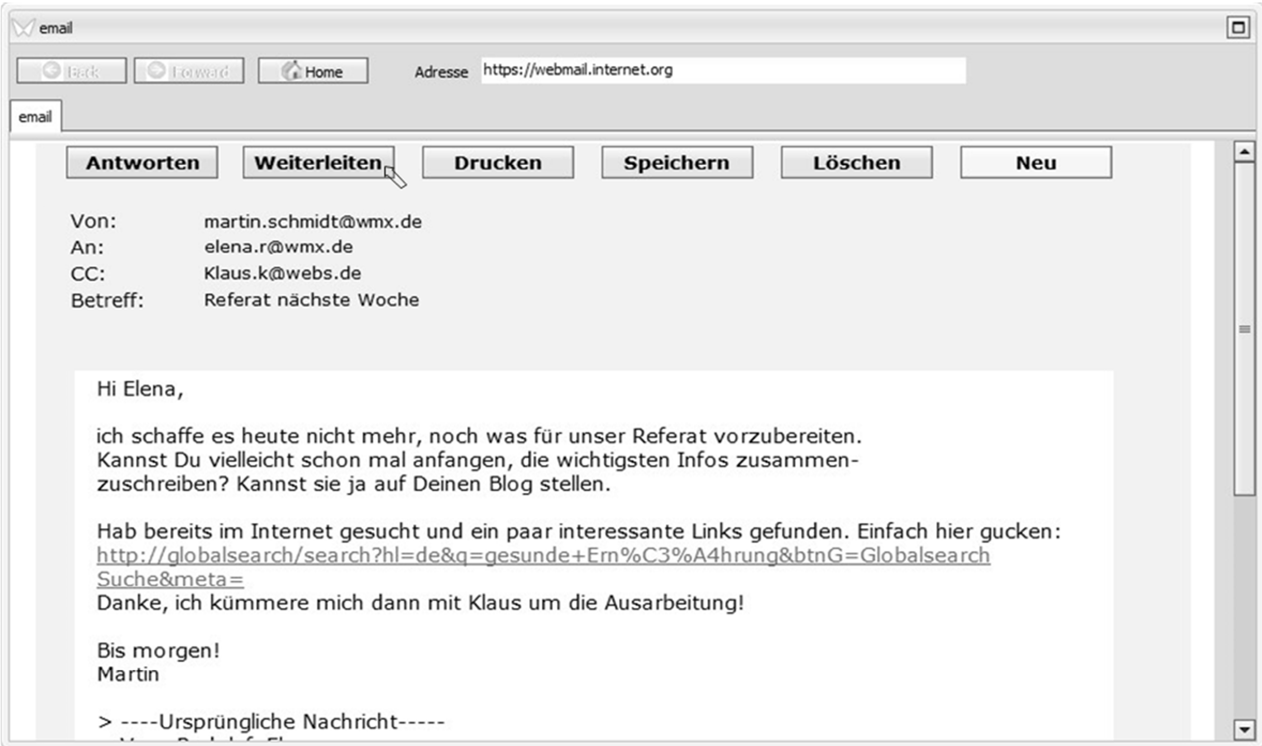


Figure 1. Two sample screenshots of BCS tasks: Screenshot (A) shows the stimulus of Task 8 “Saving as new file.” To complete it correctly, the test taker clicks on menu “File” (in German: “Datei”) and selects menu item “Save as” (in German: “Speichern unter”). Screenshot (B) shows the stimulus of Task 15 “Clicking forward button.” To complete it correctly, a click on the button “Forward” (in German: “Weiterleiten”) is required.

Table 1. Items of the BCS scale.

ID	Item	Simulated environment	Task description (instruction)	Required interactions solve the task
1	Clicking search button	Web browser	The term "New York" was typed in the search machine Global Search. Use the mouse to get the search results.	1. click button "Global – Search"
2	Scrolling	Web browser	You are at a website of the public transportation service, which has a link to the Black Forest. Find the link and click on it to enter the corresponding website.	1. scroll down 2. click link
3	Formatting	Text editor	You want to format a part of your text document in bold letters. The text part you want to format is already marked. Use the mouse to format the text part in bold letters.	1. click menu item
4	Typing	Text editor	You are shown a text, which you should typewrite errorfree in the input box below. Click ok when you are done.	1. type a text of 239 characters
5	Clicking home button	Web browser	You are on a website about butterflies and you want to go back to the start page of your browser. Use the mouse to go back to the start page.	1. click button "Home"
6	Clicking menu item	Web browser	You are on a website that shows train connections. Find the fastest train connection using the sorting function.	1. click on menu "Choose" 2. click on menu item "Duration"
7	Copy & paste	Web browser	You are shown a latin text. Use the mouse to copy the text into the input box below, then click ok.	1. highlight text 2. copy 3. paste
8	Saving as new file	Text editor	You have just updated a document and want to save it while keeping the old version. Use the mouse to save your updated text version without losing the old one.	1. click on menu "File" 2. click on menu item "Save as"
9	Searching images	Web browser	You want to search for pictures with the search engine Global Search. Use the mouse to start the picture search.	1. click button "Images"
10	Finding a string	Web browser	You are on a website of the city Münster. Use the search function to find out if the website contains information about an architect named Schlaun.	1. type the word "Schlaun" 2. click button "Search"
11	Clicking back button	Web browser	You are on a website about butterflies and you want to go back to the website that you have visited before. Use the mouse to go back to the website you have visited before.	1. click button "Back"
12	Saving file	Text editor	You want to save your text document. Use the mouse to save the text.	1. click menu item
13	Clicking reply button	E-mail client	You have received an e-mail and want to reply to it. Use the mouse to start creating a reply.	1. click button "Reply"
14	Clicking bookmark	Web browser	You are on a website and now want to switch to a website called www.nachrichten.de, for which a bookmark exists. Use the bookmark function to open the website www.nachrichten.de.	1. click button "Bookmark" 2. click on "Nachrichten.de"
15	Clicking forward button	Email client	You have received an email, which you want to forward to another person. Use the mouse to send the email to someone else.	1. click button "Forward"

2.2.2 Practical computer knowledge.

Computer knowledge was assessed with the scale *practical computer knowledge* (PRACOWI) from a German inventory for the assessment of computer literacy, computer-related attitudes and computer anxiety (Revised Computer Literacy Inventory, INCOBI-R, Richter et al., 2010; see also Naumann, Richter, & Groeben, 2001). The scale PRACOWI presents 20 everyday computer problems in the form of written scenarios. For each problem, four possible solutions are presented, one of which is correct.

The scale successfully captures the degree of practical knowledge necessary to deal with everyday computer tasks and problems. It correlates substantially with measures of computer use, such as number of desktop or WWW applications used (Naumann et al., 2001; Richter, Naumann, & Groeben, 2001; Richter et al., 2010), it differentiates between computer experts and novices (Naumann et al., 2001), and it predicts performance in everyday computer tasks. The items of the PRACOWI have been shown to be unidimensional (Richter et al., 2010).

> *Sample item:* "You want to prevent other persons from following your navigation behavior on the internet. What do you do? (1) In my computer system settings, I delete my computer's IP address. (2) In my computer system settings, I set the security settings so that my computer is invisible for others on the internet. (3) I delete all existing cookies and set the settings of my web browser to not accept new ones. (4) I delete my computer's MAC address and set the settings of my computer not to retrieve a new one."

2.2.3 Word recognition.

Students' efficiency, i.e., speed and ability of word recognition (WR) were assessed with a Lexical Decision Task (see, e.g., Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004). Students judged 30 words and 30 nonwords that appeared successively on a computer screen. Participants were asked to work as quickly as possible while avoiding errors. Each ten of the words were nouns, verbs in the infinitive form, and adjectives. They varied in length between four and eight letters and between one and three syllables. To cover a wide variety of difficulty levels, orthographic regularity and number of orthographic neighbors were varied across items. Nonwords were constructed by finding for each word another word that was equivalent in word type, length, frequency, regularity, and number of orthographic neighbors. From this second word, a nonword was constructed by switching letters. Individual differences in word recognition are reflected by both response accuracy and (log-transformed) response time.

> *Sample item (word):* "spielen" (German word for to play).
> *Sample item (nonword):* "tuckel."

2.2.4 Self-reported computer skills.

The *self-report on computer skills* included 12 items using a 4-point scale from *do well by myself* to *don't know* to indicate the confidence in doing ICT tasks. The statements referred to topics such as e-mailing, using the internet, multimedia, computer viruses, databases, and file management (for details see OECD, 2011, chapter 5).

Sample item: "To what extent are you able to do each of the following tasks on a computer? Use a spreadsheet to plot a graph:

1) Do well by myself.
2) Do with help.
3) Know but can't do.
4) Don't know."

2.2.5 Electronic reading.

Electronic reading ability was assessed by means of 67 text comprehension tasks in 13 short hypertexts. The tasks included simulated hypertext environments with text materials that are typical for electronic reading (e.g., blogs, emails, websites, online learning environments). They required the participants to use various navigation tools, e.g., clickable images or menus with links, to access textual information that was needed to answer the comprehension questions (see OECD, 2009, Annex A2; OECD, 2011, chapters 2 and 3).[1]

Sample item: Philosophers' Café – Question: "You are at the Philosophers' Café Homepage. Click on the link for Confucius. What did Confucius mean by 'Ren'?

1) Peace and prosperity.
2) Living in chaos and war.
3) The behavior of rulers.
4) Kindness to other people.
5) Living in harmony."

2.3 Procedure

The administered instruments were part of the PISA 2009 field test and a German national extension thereof. Students were tested in groups of 5–10 in computer labs at schools. First, the PISA study was introduced as an investigation of what 15-year-old students know. All

[1] Released units from the Electronic Reading Field Trial and Main Study can be viewed at http://erasq.acer.edu.au/, using "public" and "access" as username and password (released October 28, 2011).

participants were asked to complete the computer-based performance measure and then the questionnaire. The overall completion time was about 120 minutes.

2.4 Data Analysis

Psychometric properties of the BCS scale were analyzed using confirmatory factor analysis (CFA) (e.g., Bollen, 1989) and by considering coefficients related to classical test theory (CTT). CFAs were conducted to test the hypothesized unidimensional measurement models for BCS ability (measured by item response indicators), and for BCS speed (measured by item response time indicators). Moreover, CFAs were used to assess the indicators' properties and to explore the covariance structure assumed to exist between BCS ability and speed. Model parameters were estimated by means of the Mplus software (Muthén & Muthén, 1998–2010). For CFAs including continuous RT indicators, the MLR estimator being robust to nonnormality was used; to conduct CFAs including also categorical response indicators, the WLSMV estimator was used (see, e.g., Forero & Maydeu-Olivares, 2009). A model was considered to show a good overall model fit if the following criteria were met (note, values in parentheses indicate a still acceptable fit): nonsignificant χ^2 value, ratio of the χ^2 value and degrees of freedom below 2 (3), root mean square error of approximation (RMSEA) below .05 (.08), standardized root mean square residual (SRMR) below .05 (.10);weighted root mean square residual (WRMR) below .90; comparative fit index (CFI), and nonnormed fit index (NNFI) above .97 (.95) (Schermelleh-Engel, Moosbrugger, & Müller, 2003; Schweizer, 2010; for WRMR see Muthén, 1998–2004). Also, local areas of misfit were assessed by checking for patterns in the correlation residual matrix (Bollen, 1989).

CTT-related item coefficients and Cronbach's α were estimated using statistical packages of the R environment (R Development Core Team, 2009). For investigating the validity of the BCS scale, Mplus was used to estimate regression and correlation coefficients. Because participants were sampled from schools, the estimation was configured to take nonindependence of observations into account and thereby to provide correct standard errors (Mplus option TYPE = COMPLEX).

3. Results

3.1 CFA Modeling of the BCS Tasks

Measurement models were tested to test the unidimensionality of BCS speed and ability as claimed in Hypothesis 1. Unidimensional CFAs were conducted for responses and response times first separately and then jointly to explore the covariance structure of BCS ability and speed. Item 2 was dropped from the beginning because it did not show any variance in responses.

Results for the initial RT-CFA model showed that item 4 did not load significantly on the latent speed factor (standardized solution: $\lambda_{4Speed} = .08$, $z = 1.12$, $p = .27$). However, it was not dropped as it may measure BCS ability well. According the model fit criteria as described above, the fit of the model was very good. This unidimensional model was compared to a three-dimensional model assuming environment-specific speed factors. As expected, the obtained fit for the less restricted model was also very good. Most importantly, however, the unidimensional model did not fit the data worse than the three-dimensional model as indicated by the Wald test, $\chi^2(3, N = 315) = .87$, $p = .83$, suggesting retention of the more parsimonious unidimensional model.

The initially tested R-CFA model including all items except for item 2 did not fit the data well, $\chi^2(77, N = 315) = 127.83$, $p < .01$, RMSEA = .046, WRMR = 1.22, CFI = .61, NNFI = .54). The inspection of residuals revealed that residuals were greatest for item 3. By excluding item 3, the model fit became acceptable. Again, the final unidimensional model was compared with a model assuming three environment-specific ability factors. The fit for the less restricted model was also acceptable. Once again, however, the unidimensional model did not fit the data worse than the three-dimensional model as indicated by the Wald test, $\chi^2(3, N = 315) = .81$, $p = .85$.

Based on these results supporting the hypothesis of unidimensionality of both BCS speed and ability, a joint RTR-CFA model was tested including all but the dropped items 2 and 3. The overall fit of the model was not entirely acceptable, $\chi^2(298, N = 315) = 377.65$, $p < .01$, RMSEA = .029, WRMR = 1.03, CFI = .84, NNFI = .83. Modification indices suggested to free the item-specific residual correlations between response and response time indicator of item 4 and item 7, respectively. The fit of the modified model was acceptable. Results showed that the variances of both BCS speed and ability were significant, Var(Speed) = .08 ($z = 3.79$, $p < .01$), and Var(Ability) = .36 ($z = 2.04$, $p = .04$), as well as their correlation of Cor(Speed, Ability) = .56 (z = 6.95, p < .01). The first result indicates systematic variance between individuals in their BCS speed and ability. The latter result indicates that, as expected, participants with higher levels of ability also tend to work at a higher speed.

Table 2. Model parameters of the final CFA model for responses and response times in BCS items and corresponding CTT item coefficients (items 2 and 3 were not included in the final CFA model).

		CFA model parameters				CTT item coefficients		
		Response model		Response time model		Response data		Response time data[a]
ID	Item	Threshold	Loading (standardized)	Intercept	Loading (standardized)	*M*(R) Difficulty	*Cor*(R_i,R_t) Discrimination	*Cor*(RT_i,RT_t) Time-related discrimination
1	Clicking search button	–1.87	.60	8.74	.48	.97	.27	.53
2	Scrolling	–	–	–	–	–	–	–
3	Formatting	–	–	–	–	–	–	–
4	Typing	1.09	.43	11.92	.12	.14	.50	.24
5	Clicking home button	–0.03	.30	8.90	.62	.51	.43	.62
6	Clicking menu item	–1.18	.41	10.13	.60	.88	.33	.59
7	Copy & paste	.48	.34	10.13	.57	.32	.52	.62
8	Saving as new file	.58	.33	9.48	.60	.28	.57	.64
9	Searching images	–1.52	.14	8.57	.69	.94	.03	.68
10	Finding a string	.75	.49	9.43	.29	.23	.47	.41
11	Clicking back button	–1.05	.44	8.38	.63	.85	.25	.66
12	Saving file	–1.25	.45	8.80	.60	.89	.25	.65
13	Clicking reply button	–1.52	.24	8.78	.54	.94	.26	.59
14	Clicking bookmark	–.99	.61	9.63	.65	.84	.38	.66
15	Clicking forward button	–1.84	.72	8.69	.55	.97	.11	.58

Notes. R = response, RT = log-transformed response time, *Cor*(R_i,R_t) = point biserial correlation between item score and total score, *Cor*(RT_i,RT_t) = Pearson correlation between item response time and total response time. [a]The values for *M*(RT) per item are not presented as they correspond to the intercepts of the response time CFA model (expected value of RT_i).

Table 2 shows the estimated CFA model parameters of the final joint measurement model including items 1 and 4 to 15. Most of the thresholds are negative, indicating that the tasks are easy, as it can be expected for tasks measuring basic ICT skills. In the RT model, the intercept parameter reflects the required time corresponding to the average RT; the factor loading indicates how well an item distinguishes between participants' speed levels. Although items 4 and 9 showed low factor loadings on BCS speed and BCS ability, respectively, they were not dropped because of good measurement properties related to BCS ability and BCS speed, respectively. Table 2 also presents item coefficients related to CTT, i.e., for response data, item difficulty *M*(R) and discrimination *Cor*(R_i,R_t), and similarly for response time data the RT-related discrimination *Cor*(RT_i,RT_t). Reliability analysis revealed Cronbach's α values of .84 for BCS speed and of .70 for BCS ability.

In sum, the part of Hypothesis 1 proposing unidimensionality of speed and ability was supported for the revised scale. This means the data support the notion that for test takers' accuracy in completing the BCS tasks one latent variable is sufficient to explain individual differences in accuracy. Thus, as regards the materials used in the present study, there appears to be one BCS ability rather than multiple abilities relating to multiple environments. The same claim can be made for the observed response times: Individual differences are again best explained by the assumption that there is one speed of BCS task completion rather than multiple speeds that are distinguished along the lines of different environments.

BCS ability and BCS speed were also substantially related to one another, sharing about one fourth of their variance. This indicated that test takers with higher levels of ability also work at a higher level of speed. However, the correlation of BCS speed and ability was far less than one, meaning that, while BCS speed and ability are related, they are by no means the same – evidence for the notion that it is necessary to measure them both. In other words: Measuring BCS speed gives information about a test taker that is not available if only BCS ability is measured, and vice versa.

3.2 Evidence for the Convergent and Discriminant Validity of the BCS Scale

As a preparatory step of the following validation, CFAs were conducted for scales assessing practical computer knowledge, word recognition (WR, both WR speed and ability), self-reported computer skills, and electronic reading ability. CFAs served to test the assumed unidimensionality of these scales and to estimate factor scores. Table 3 shows the distribution parameters, reliability coefficients, and Pearson correlations for factor scores of all measures. Because the electronic reading items were delivered in a booklet design, Cronbach's α could not be computed. We therefore present the IRT-based reliability (see Wu, Adams, Wilson, & Haldane, 2007).

Table 3. Distribution parameters, scale reliability, and Pearson correlations of Basic Computer Skills (BCS), Practical Computer Knowledge, Word Recognition (WR), Self-Reported Computer Skills, and Electronic Reading.

Measure	*M*	*SD*	Skewness	Kurtosis	Cronbach's α	Pearson correlation 1	2	3	4	5	6
1 BCS – Ability	–.03	0.46	–.26	–.24	.70						
2 BCS – Speed[a]	.01	0.26	.32	.12	.81	.74					
3 Practical Computer Knowledge	.01	0.45	.36	–.05	.81	.60	.47				
4 WR – Ability	.00	0.12	–1.89	3.48	.91	.32	.23	.24			
5 WR – Speed[a]	.00	0.23	.18	.54	.96	.29	.36	.16	–.03		
6 Self-reported Computer Skills	–.03	0.57	–.41	.59	.86	.26	.18	.35	.04	.13	
7 Electronic Reading	.04	1.70	–.27	–.56	.71[b]	.60	.54	.54	.44	.30	.16

Notes. [a]Factor scores from CFA actually reflect slowness; for the computation of correlations, factor scores were multiplied by (–1) so that all measures have a positive orientation. [b]IRT-based reliability.

3.2.1 Results on Hypothesis 2.1: Predictive validity of practical computer knowledge and word recognition.

2.1a: The regressions of BCS speed and BCS ability on practical computer knowledge revealed, as hypothesized, significant (standardized) regression coefficients of $b = .47$ ($z = 9.61$, $p < .01$) for BCS speed and of $b = .60$ ($z = 17.15$, $p < .01$) for BCS ability.

2.1b: Also when regressing BCS speed on WR speed and BCS ability on WR ability, significant regression coefficients of $b = .36$ ($z = 4.01$, $p < .01$) for BCS speed and of $b = .32$ ($z = 5.05$, $p < .01$) for BCS ability could be shown, albeit smaller than those found for practical computer knowledge.

2.1c: Multiple regression of BCS speed on practical computer knowledge and WR revealed significant associations of BCS speed with both practical computer knowledge of $b = .40$ ($z = 8.32$, $p < .01$), and with WR speed of $b = .30$ ($z = 3.71$, $p < .01$) when accounting for the respective other predictor. The proportion of explained variance was $R^2 = .29$ (for practical computer knowledge $\Delta R^2 = .16$, for WR speed $\Delta R^2 = .07$). When BCS ability was regressed on practical computer knowledge and WR ability, again significant associations were obtained for both practical computer knowledge of $b = .55$ ($z = 12.66$, $p < .01$), and for WR ability of $b = .18$ ($z = 3.14$, $p < .01$). For BCS ability the amount of explained variance of $R^2 = .38$ was even greater than for BCS speed (for practical computer knowledge $\Delta R^2 = .28$, for

WR ability $\Delta R^2 = .02$). The fact that associations of practical computer knowledge and BCS remained after including WR as a predictor in the model indicated that the prediction of BCS ability and speed from practical computer knowledge is not an artefact driven by the fact that reading is involved in solving both BCS tasks and practical computer knowledge tasks. It also indicated that rapid access to the meaning of words is predictive of task success in digital environments requiring the decoding of verbal information, e.g., as displayed on buttons. This is in line with the theoretical notion that at least basic decoding skills are required to successfully cope with even low level computer tasks.

3.2.2 Results on Hypothesis 2.2: Relationship to self-reported computer skills.

As expected, the correlation between self-reported computer skills and BCS speed of $r = .18$ was small but significant ($z = 2.03$, $p = .04$). Similarly, the obtained correlation of self-reported computer skills with BCS ability of $r = .26$ was moderate and significant ($z = 4.19$, $p < .01$). This indicated that students' perceptions of their ICT skills were related to their actual skills as measured by BCS speed and ability, though these relations were weaker than the relationships of an objective test of practical computer knowledge with BCS speed and ability.

3.2.3 Results on Hypothesis 2.3: Relationship to gender.

When BCS speed was regressed on gender, a significant regression coefficient of $b = .21$ ($z = 4.11$, $p < .01$) was found. When BCS ability was regressed on gender, a similar regression coefficient of $b = .24$ ($z = 5.94$, $p < .01$) was obtained. The proportions of explained variance for BCS speed was $R^2 = .04$, and for BCS ability it was $R^2 = .06$. As expected, the positive regression coefficients indicated a better performance of males in BCS tasks than females; however, the amount of explained variance by gender was only small.

3.2.4 Results on Hypothesis 2.4: Prediction of electronic reading ability.

Finally, the predictive validity of BCS with electronic reading was investigated. As expected, electronic reading was predicted by BCS ability, $b = .45$ ($z = 5.23$, $p < .01$), and BCS speed, $b = .21$ ($z = 2.18$, $p = .03$). The regression coefficients indicate that BCS ability and speed independently predict electronic reading. The proportion of variance in electronic reading skill explained by BCS ability and speed was $R^2 = .38$. This result indicates that electronic reading, which in itself constitutes an important aspect of computer use in an information society, is largely predicted by BCS speed and ability. This also agrees with current models of problem solving in the context of ICT, assuming basic computer skills are a prerequisite of higher-order problem-solving activities (Brand-Gruwel et al., 2009).

Taken together, the hypotheses pertaining to the internal structure of BCS and its association with other cognitive as well as demographic variables were confirmed. As expected, BCS ability and speed proved to be unidimensional and separable dimensions of BCS. Regarding their assumed cognitive basis, BCS ability and speed had substantial correlations and partial correlations with practical computer knowledge, as well as word recognition. There were lower but significant correlations with self-assessed computer skills. Also, as suggested by previous findings, males had slightly higher levels of BCS ability and speed than females. Finally, as hypothesized, BCS ability and speed were shown to be underlying components of electronic reading ability.

4. Discussion and Conclusions

The present study served to develop a BCS scale allowing for an assessment of the participant's ability as well as speed in completing basic computer tasks. The task design was based on a definition that conceptualizes BCS as ability and speed of accessing, collecting, and providing information by performing basic actions on computer graphical user interfaces. A computer-based assessment approach was selected to obtain a more valid measurement by means of interactive and simulation-based task types as well as to collect response time data for measurement models taking both responses and response times into account. For validation, we considered various constructs related to computer skills and measurement approaches.

Empirically, unidimensionality in both BCS speed and BCS ability could be assumed after dropping two ill-fitting tasks. It is important to note that the present results need to be replicated in independent samples in future studies. If this result holds up, it would confirm the theoretical assumption that BCS represent generic skills that are not specific to particular environments but refer to task demands common to different software applications.

BCS speed and BCS ability were substantially related, indicating that accurate participants also tend to be fast. However, the amount of unique variance suggests that both person parameters carry specific information, i.e., in a group of participants with a given ability level there will still be variability in the level of speed and vice versa. The empirical separability of BCS ability and speed supports the theoretical notion that both ability and speed are specific constituents of BCS. Practically speaking, this implies that the assessing and reporting of BCS should take both speed and ability into account (e.g., Sandene et al., 2005).

As it stands the scale appears to provide a valid measure of BCS, according to a number of relations to other measures, both in terms of convergent and discriminant validity. First, both BCS speed and ability show strong correlations with knowledge on the solution of higher order ICT tasks (Practical Computer Knowledge). This is in line with models such as the IPS-I model of problem solving on the internet, where basic computer skills are deemed a requirement for the solution of higher-order tasks. Second, relations with basic reading skills are in line with the fact that the decoding of small text segments was required for the basic ICT tasks appearing on the test, starting with reading the task and ending with, say, pressing a particular button with a textual label. Indeed, while models like the IPS-I assume reading skills to be a basic requirement of problem solving on the internet, we would claim that at least basic reading is also a prerequisite for solving basic ICT tasks. However, while there were correlations of BCS and basic reading, these were significantly lower than the association of BCS with practical computer knowledge. Finally, BCS speed and ability were found to predict performance on a measure of an essential 21st-century skill – comprehension of online text – including the selection and integration of various text segments from a complex information space. As expected, these associations were again stronger than the association with pure word-decoding skill. All in all, the pattern of associations of BCS with other measures warrant the conclusion that the scale proposed here provides a valid measure of a person's skill in carrying out basic operations in using the computer to access, collect, and provide information.

Hypotheses on validity were put forward for both BCS speed and ability. The most noticeable empirical difference was that practical computer knowledge (as well as electronic reading ability) showed stronger associations with BCS ability than with BCS speed. One explanation would be that ability-related measures (here: BCS ability) generally show higher commonalities with other ability-related measures (here: practical computer knowledge) than with speed-related measures (here: BCS speed) due to a general underlying cognitive ability component (see Carroll, 1993). However, the low correlation of .32 between BCS ability and WR ability as another ability parameter suggests that the strong association of BCS ability with practical computer knowledge cannot be explained solely by an unspecified underlying cognitive ability. Rather, according to Hypothesis 2.1, practical computer knowledge seems to be a domain-specific condition for BCS.

Taken together, the results obtained suggest that the BCS scale is a reliable and valid means of assessing a person’s basic computer skills, both in educational and research contexts. Possible applications of the scale in applied settings might include the tailoring of computer-based instruction providing students with more study time who, for example, because of limited access to computers, already struggle with very basic operations of computer-based information access. In research settings, the scale described here provides researchers with a covariate for disentangling different sources of variance in performance of more complex ICT tasks. For instance, as shown by results from PISA 2009 from around the world, large proportions of students in a number of countries have problems accessing information in hypertext environments. As a consequence, they receive only poor scores on a digital reading test (see OECD, 2011, chapter 3). Including a measure of basic ICT skills in such a study might help to determine whether these students lack BCS or whether it is more a matter of metacognitive strategies when evaluating the contents of a document that prevent them from gathering the information they need.

For the further development of the BCS scale, some extensions and modifications of the present version need to be considered. First, to obtain a more systematic construct representation, the BCS scale should be balanced with respect to simulated environments and information skills, i.e., tasks should be added that simulate an e-mail client, text editor or processor, and file manager, and some more tasks should require participants to collect information and to provide it to others. Items 3, 4, and 9 should be replaced as they showed item misfit or low discrimination with respect to BCS speed and BCS ability, respectively. One limitation of the present scale is the restricted range of item difficulties. All items are of low and medium difficulty, i.e., participants with high levels of BCS ability cannot be discriminated. To increase the benefit from the BCS ability parameter, some more difficult tasks should be added to the scale by increasing task complexity, while maintaining the scale capturing basic skills of accessing, collecting, and sharing computer-based information. Future studies should also aim at supplementing the present findings with the investigation of how BCS ability and speed are related to other ICT literacy competencies or performance in other ICT tasks such as creating information using word processors or graphics programs, or managing information using, e.g., databases.

Acknowledgments

The authors are grateful to Alexander During, Rachel Ghebrehawariat, Thomas Martens, Heiko Rölke, Britta Upsing, and Merle Steinwascher for their support. The authors also want to thank two anonymous reviewers for their helpful remarks on an earlier version of this paper.

References

Anderson, J.R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*, 369–406.

Anderson, J.R., Matessa, M., & Lebiere, C. (1997). ACT-R: A theory of higher level cognition and its relation to visual attention. *Human-Computer Interaction, 12*, 439–462.

Ballantine, J.A., Larres, P.M., & Oyelere, P. (2007). Computer usage and the validity of self-assessed computer competence among first-year business students. *Computers and Education, 49*, 976–990.

Balota, D.A., Cortese, M.J., Sergent-Marschall, S.D., Spieler, D.H., & Yap, M. (2004). Visual word recognition of single syllable words. *Journal of Experimental Psychology: General, 133*, 283–316.

Bollen, K.A. (1989). *Structural equations with latent variables.* New York, NY: Wiley.

Brand-Gruwel, S., Wopereis, I., & Walraven, A. (2009). A descriptive model of information problem solving while using internet. *Computers and Education, 53*, 1207–1217.

Carroll, J.B. (1993). *Human cognitive abilities*. New York, NY: Cambridge University Press.

Colley, A., & Comber, C. (2003). Age and gender differences in computer use and attitudes among secondary school students: What has changed? *Educational Research, 45*, 155–165.

Educational Testing Service. (2008). *iSkills – Information and Communication Technology Literacy Test.* Retrieved from http://www.ets.org/iskills/about

Forero, C.G., & Maydeu-Olivares, A. (2009). Estimation of IRT graded models for rating data: Limited vs. full information methods. *Psychological Methods, 14*, 275–299.

Goldhammer, F., & Klein Entink, R.H. (2011). Speed of reasoning and its relation to reasoning ability. *Intelligence, 39*, 108–119.

Hakkarainen, K., Ilomäki, L., Lipponen, L., Muukkonen, H., Rahikainen, M., Tuominen, T., Lehtinen, E. (2000). Students skills and practices of using ICT: Results of a national assessment in Finland. *Computers and Education, 34*, 103–117.

Ilomäki, L., & Rantanen, P. (2007). Intensive use of ICT in school: Developing differences in students' ICT expertise. *Computers and Education, 46*, 119–136.

Imhof, M., Vollmeyer, R., & Beierlein, C. (2007). Computer use and the gender gap: The issue of access, use, motivation, and performance. *Computers in Human Behavior, 23,* 2823–2837. International ICT Literacy Panel. (2002). *Digital transformation: A framework for ICT literacy (A report of the International ICT Literacy Panel).* Princeton, NJ: Educational Testing Service. Retrieved from http://www.ets.org/Media/Research/pdf/ictreport.pdf

Katz, I.R., & Macklin, A.S. (2007). Information and communication technology (ICT) literacy: Integration and assessment in higher education. *Journal of Systemics, Cybernetics and Informatics, 5*, 50–55.

Keith, N., & Frese, M. (2005). Self-regulation in error management training: Emotion control and metacognition as mediators of performance effects. *Journal of Applied Psychology, 90*, 677–691.

Keith, N., Richter, T., & Naumann, J. (2010). Active/exploratory training promotes transfer even in learners with low motivation and cognitive ability. *Applied Psychology: An International Review, 59*, 97–123.

Kim, J.H., Jung, S.Y., & Lee, W.G. (2008). Design of contents for ICT literacy in-service training of teachers in Korea. *Computers and Education, 51*, 1683–1706.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York, NY: Cambridge University Press.

Klieme, E., Hartig, J., & Rauch, D. (2008). The concept of competence in educational contexts. In J. Hartig, E. Klieme & D. Leutner (Eds.), *Assessment of competencies in educational contexts* (pp. 3–22). Göttingen: Hogrefe.

Kuhlemeier, H., & Hemker, B. (2007). The impact of computer use at home on students' internet skills. *Computers and Education, 49*, 460–480.

Lennon, M., Kirsch, I., Von Davier, M., Wagner, M., & Yamamoto, K. (2003). *Feasibility study for the ICT Literacy Assessment: Report to network A*. Retrieved from http://www.oecd.org/dataoecd/35/13/33699866.pdf

Leu, D.J., Kinzer, C.K., Coiro, J.L., & Cammack, D.W. (2004). Toward a theory of new literacies emerging from the internet and other information and communication technologies. In N.J. Unrau & R.B. Ruddell (Eds.), *Theoretical models and processes of reading* (5th ed., pp. 1570–1613). Newark, NJ: International Reading Association.

Mabe, P.A., & West, S.W. (1982). Validity of self-evaluation of ability: Review and meta-analysis. *Journal of Applied Psychology, 67*, 280–296.

Markauskaite, L. (2007). Exploring the structure of trainee teachers' ICT literacy: The main components of, and relationships between, general cognitive and technical capabilities. *Educational Technology Research and Development, 55*, 547–572.

Mayer, R.E. (2003). What causes individual differences in cognitive performance? In R.J. Sternberg & E.L. Grigorenko (Eds.), *The psychology of abilities, competencies, and expertise* (pp. 263–74). New York, NY: Cambridge University Press.

Metzger, M.J., Flanagin, A.J., & Zwarun, L. (2003). College student web use, perceptions of information credibility, and verification behavior. *Computers and Education, 41*, 271–290.

Muthén, B.O. (1998–2004). *Mplus technical appendices*. Los Angeles, CA: Muthén & Muthén.

Muthén, L.K., & Muthén, B. (1998–2010). *Mplus user's guide. Version 6*. Los Angeles, CA: Muthén & Muthén.

Naumann, J., Richter, T., & Groeben, N. (2001). Validierung des Inventars zur Computerbildung (INCOBI) anhand eines Vergleichs von Anwendungsexperten und Anwendungsnovizen [Validation of the Computer Literacy Inventory (INCOBI) by a comparison of expert and novice computer users]. *Zeitschrift für Pädagogische Psychologie/German Journal of Educational Psychology, 15,* 219–232.

OECD. (2005). *Bildungspolitische Analyse 2004* [Education policy analysis 2004]. Paris: Author.

OECD. (2009). *PISA 2009 assessment framework*. Paris: Author.

OECD. (2011). *PISA 2009 results Vol. VI. Students on line: Reading and using digital information*. Paris: Author.

Parshall, C.G., Spray, J.A., Kalohn, J.C., & Davey, T. (2002). *Practical considerations in computer-based testing*. New York, NY: Springer.

Potosky, D. (2007). The Internet Knowledge (iKnow) measure. *Computers in Human Behavior, 23*, 2670–2777.

Poynton, T.A. (2005). Computer literacy across the lifespan: A review with implications for educators. *Computers in Human Behavior, 21*, 861–872.

R Development Core Team. (2009). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Available at http://www.R-pro-ject.org

Richter, T., Naumann, J., & Groeben, N. (2001). Das Inventar zur Computerbildung (INCOBI): Ein Instrument zur Erfassung von Computer Literacy und computerbezogenen Einstellungen bei Studierenden der Geistes- und Sozialwissenschaften [The Computer Literacy Inventory: An instrument for the assessment of computer literacy and computer related attitudes in students of the humanities and social sciences]. *Psychologie in Erziehung und Unterricht, 48*, 1–13.

Richter, T., Naumann, J., & Horz, H. (2010). Eine revidierte Fassung des Inventars zur Computerbildung (INCOBI-R) [A revised version of the Computer Literacy Inventory]. *Zeitschrift für Pädagogische Psychologie/German Journal of Educational Psychology, 24*, 23–37.

Russell, M., Goldberg, A., & O'Connor, K. (2003). Computer-based testing and validity: A look back into the future. *Assessment in Education: Principles, Policy and Practice, 10*, 279–293.

Sandene, B., Bennett, R.E., Braswell, J., & Oranje, A. (2005). Online assessment in mathematics. In B. Sandene, N. Horkay, R.E. Bennett, N. Allen, J. Braswell, B. Kaplan, & A. Oranje (Eds.),*Online assessment in mathematics and writing: Reports from the NAEP Technology-based Assessment Project (NCES 2005–457)* (pp. 1–68). Washington, DC: US Department of Education, National Center for Education Statistics.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Test of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research – Online, 8*, 23–74. Available at http://www.mpr-online.de/

Schumacher, P., & Morahan-Martin, J. (2001). Gender, internet and computer attitudes and experiences. *Computers in Human Behavior, 17*, 95–110.

Schweizer, K. (2010). Some guidelines concerning the modeling of traits and abilities in test construction. *European Journal of Psychological Assessment, 26*, 1–2.

Tsai, M.-J. (2002). Do male and female students often perform better than female students when learning computers? A study of Taiwanese eight graders' computer education through strategic and cooperative learning. *Journal of Educational and Computing Research, 26*(1), 67–85.

van der Linden, W.J. (2009). Conceptual issues in response-time modeling. *Journal of Educational Measurement, 46*, 247–272.

Vekiri, I., & Chronaki, A. (2008). Gender issues in technology use: Perceived social support, computer self-efficacy and value beliefs, and computer use beyond school. *Computers and Education, 51*, 1392–1404.

Wecker, C., Kohnle, C., & Fischer, F. (2007). Computer literacy and inquiry learning: When geeks learn less. *Journal of Computer-Assisted learning, 23*, 133–144.

Weinert, S., Artelt, C., Prenzel, M., Senkbeil, M., Ehmke, T., & Carstensen, C.H. (2011). Development of competencies across the life span. *Zeitschrift für Erziehungswissenschaft,14*, 67–86.

Wu, M.L., Adams, R.J., Wilson, M.R., & Haldane, S. (2007). *ACER ConQuest 2.0: General item response modeling software*. Camberwell, Australia: Australian Council for Educational Research.

Paper 2

Publication Note

Keßel, Y., Goldhammer, F., & Kröhne, U. (under review, *Interacting with Computers*). Evaluating the Credibility of Online Information: The Influence of Basic Computer Skills, Word Recognition and Reasoning.

Evaluating the Credibility of Online Information: The influence of basic computer skills, word recognition and reasoning

A newly developed test measuring the ability to evaluate the credibility of online information is introduced. The evaluation of online information is conceptualized to include the judgement of structural features (e.g., domain name) and message features (e.g., information accuracy). The ability of evaluating online information (Evaluation of Structural and Message Features, ESMF) was shown to comprehend the abilities of evaluating primarily structural features (Evaluation of Structural Features, ESF) and evaluating primarily message features (Evaluation of Message Features, EMF). The latent structure of ESMF and its relationship with basic computer skills, word recognition and reasoning were analysed in a sample of N=205 high-school students in grade 12. Word recognition and reasoning revealed to be prerequisite skills for the evaluation of online information, suggesting that trainings for Web users might benefit from their improvement. Differences in the relations of the covariates with ESMF, ESF and EMF are discussed.

1. Introduction

Going online and using search engines to satisfy information needs has become an ubiquitous part of our daily life. Through the World Wide Web (Web) more information from more sources is available and more easily accessible than ever before (Metzger, 2007; Go, You, Jung, & Shim, 2016). Using search engines is quick, easy and cost-efficient. But despite their obvious advantages, they also pose a challenge to the user (Thomm & Bromme, 2012). Online information mostly does not underlie any editorial review and it may be easily altered, plagiarized, misrepresented or created anonymously under false pretenses (Britt & Aglinskas, 2002; Gerjets, Kammerer, & Werner, 2011; Metzger, Flanagin, Eyal, Lemus, & McCann, 2003; Rieh, 2002). Especially the introduction of Web 2.0 has led to an enormous increase of information suppliers (Lucassen & Schraagen, 2011). Therefore the ability to critically evaluate the credibility of online information is of great importance (Brand-Gruwel & Stadtler, 2011; Kammerer & Gerjets, 2012, Go et al., 2016) and it is "particularly necessary when searching for complex information…" (Salméron, Kammerer, & Garcia-Carrión, 2013, p. 2161). It may be crucial for a student's educational outcomes or adults' occupational success because misleading and incorrect online information used as the basis for decisions in critical domains can result in serious consequences (Gerjets et al., 2011; Schwarz & Morris, 2011). The present paper focuses on one major quality aspect of information, namely credibility reflecting the degree to which the information is trustworthy and reliable (see also Fogg & Tseng, 1999; Metzger et al., 2003).

Research has shown that people differ substantially in the ability to evaluate online search results with respect to their credibility and that many have problems in appropriately evaluating online information when searching the Web (Brand-Gruwel, Wopereis, & Vermetten, 2005; Gerjets & Hellenthal-Schorr, 2008; Hargittai, Fullerton, Menchen-Trevino, & Thomas, 2010; Meola, 2004). Even students, being frequent users of the Web, display major imperfections evaluating "search engine result pages" (SERPs) and Web sites (Brand-Gruwel, Kammerer, van Meeuwen, & van Gog, 2017; Lorenzen, 2001; MaKinster, Beghetto, & Plucker, 2002; Wallace, Kupperman, Krajcik, & Soloway, 2000; Walraven, Brand-Gruwel, & Boshuizen, 2009). For instance, Gerjets et al. (2011) have shown that individuals predominantly access the first few Web pages presented on a SERP neglecting the fact that even the top search results might not be credible information sources and one-sided or commercially biased due to search engine optimization business (Lewandowski, 2011, 2013).

The overall goal of this study is to investigate the ability of evaluating the credibility of online information from an individual differences perspective. Here, online information refers to any information encountered online by using a search engine including search engine result pages (SERPs) and any form of Web sites. A newly developed test measuring this ability is introduced. Furthermore the relations of selected cognitive components to this ability are explored. Investigating the evaluation of online information and its cognitive basis will further clarify the theoretical and empirical specificity of this ability and its anchoring in various cognitive constituents, respectively. This will shed some light on the sources of individual differences in the ability of evaluating online information that has become highly important with the progression of information and communication technologies and yet has only been understood partly.

2. Theoretical Background

2.1 The Web Search Process

The evaluation of online information is conceived as part of the Web search process describing how people search for information on the Web and which evaluation criteria are incorporated in this Web search process. This multi-stage process has been described by several models and can also be considered to be a special instance of help seeking or problem-solving process, in which a search engine and the Web provide help to deal with an informational impasse (cf. Mercier & Frederiksen, 2008; Brand-Gruwel et al., 2017). It is summarized in five important stages by Gerjets et al. (2011): (a) An information deficiency is identified and a search goal is defined. (b) To access the information in demand, a search engine (e.g., Google) is selected, a search terms is entered and the search is started. In the following, a so called ‘‘search engine result page’’ (SERP) with a list of search result links is returned to the user. (c) The list of search results is scanned and the links are evaluated with regard to their significance for the search goal. Links are selected for further inspection. At this time the evaluation of search results is based on sparse information about the corresponding Web sites and the information they contain (e.g., titles, text fragments, domain names and suffixes). One major (observable) action in the information search process is the clicking on a link and entering the corresponding Web site. (d) After entering a selected Web site, its information is scanned, evaluated with regard to its relevance for the search goal, and in case of relevance information is extracted for further processing. (e) The information from different Web sites is compared, evaluated with regard to its credibility and integrated towards a solution of the information problem (p. 221).

The present study addresses the evaluation of links on SERPs and corresponding Web sites as well as the comparison of the different links and Web sites with respect to their credibility. Therefore, in this study the focus will be on the last three stages of the Web search process. In the following we elaborate on the properties of a Web site which are assumed to reflect the credibility of online information, and, therefore, may affect individual credibility judgments when evaluating online information.

2.2 Credibility Judgment as Information Processing

2.2.1 Information features.

With the rise of user-generated content on the Web and the urgency to properly evaluate online information, the question of which skills are needed to successfully cope with this task has been topic of research for some time now and produced several models explaining information evaluation behaviour online. According to Prominence-Interpretation-Theory (P-I-Theory, Fogg, 2002) internet users first notice an information feature (Prominence) and then make a judgment about it (Interpretation). Both steps are necessary conditions for a credibility assessment. This process of noticing and interpreting information features will typically happen numerous times during the evaluation of online information, with new aspects on a SERP or Web site being noticed and interpreted (Fogg, 2002). Thus, evaluating the credibility of online information refers to the judgment of many different information characteristics, which can be classified as message features reflecting the content of information (e.g. statistics, testimonials) and structural features reflecting its presentation (e.g. navigation tools such as site maps, domain names and suffixes) (Burkell, 2004; Hong, 2006; Lucassen & Schraagen, 2011, Rains & Karmikel, 2009; Wathen & Burkell, 2002). In the remainder of this paper we refer to the *evaluation of message features (EMF)* when investigating the message content by reading a text, while we refer to the *evaluation of structural features (ESF)* when investigating accompanying information found in its presentation.

As for message features and credibility, an objectively verifiable relationship is also assumed for structural features and credibility. Based on this assumption, a Web site is assumed to be a credible source of information if the message features concerning text quality are fulfilled, whereas the presumed structural features for high quality Web sites allow this conclusion in advance with high probability.

2.2.2 The evaluation of SERPs and web sites.

In principal, message and structural features can be encountered both on SERPs and Web sites because both comprise text (message content) as well as accompanying information relating to its presentation. However, it can be assumed that the evaluation of SERPs differs from the evaluation of Web sites. Links on SERPs provide only little information about their destinations as they are confined to a very small amount of message content (Wirth, Böcking, Karnowski, & von Pape, 2007). On SERPs prominent credibility features are structural features, whereas for the evaluation of Web sites message features are more eminent (Gerjets et al., 2011). The evaluation of links on a SERP is characterized by high uncertainty (due to minimal content information) and high subjective task complexity fostering heuristic information processing (Wirth et al., 2007). As SERPs contain only minimal content descriptions the user must infer message features from structural features; hence structural features are used to predict message features based on heuristic rules. For instance, by looking at a domain name on a SERP, inferences about the message content (e.g. objectivity) existent on the corresponding Web site can be drawn. This step can be compared to the concept of a predictive judgment according to Rieh (2002, p. 146) as it “guides a decision about what kind of action the user is going to take given multiple choices (alternatives)”. Inferences about unknown aspects have to be made. Following Gigerenzer and Goldstein (1996) this type of inference is made based on one-reason decision making. That means instead of “taking the time to survey all possible alternatives” (Gigerenzer & Goldstein, 1996, p. 651) a decision is made by choosing the first alternative, which satisfies the aspiration level. If the information in a link found on a SERP is not evaluated as credible by the user, he or she will probably not make the effort to enter the corresponding Web site (predictive judgment) allowing a thorough and more systematic evaluation of all available structural and message features (ESMF), which according to Rieh (2002) would be an evaluative judgment. The dual-processing model by Metzger (2007), based on the dual-processing theory of Chaiken (1980), predicts whether a heuristic or systematic information processing is done. According to this model, heuristic evaluation means using superficial cues to judge credibility. In line with Metzger (2007) and Fogg's P-I-Theory (2002) we believe that the evaluation of SERPs is mainly heuristic because it rather allows the evaluation of structural information features as they are the more prominent information features in this case. Thus, when selecting search results from a SERP, we assume that it is of primary importance to make credibility judgments about structural features and be able to carry out heuristic information processing, whereas in other cases, for instance, when inspecting entries in an internet forum it might be essential to effectively evaluate available message features (EMF) by thoroughly reading the presented text. We assume that the ability of evaluating online information based on structural and message features (Evaluation of Structural and Message Features, ESMF) underlies two separable evaluation skills, namely the evaluation of online search results based on structural features only (Evaluation of Structural Features, ESF) and the evaluation of online search results based on message features only (Evaluation of Message Features, EMF).

2.3 Individual Differences in Evaluating Online Information

Internet users are assumed to differ in their ability to distinguish between high and low quality Web sites. First, when evaluating online information one needs to know and identify which features are important to look at. Lucassen and Schraagen (2010) showed that there is a difference between high-school students and college students in judging the importance of references when evaluating credibility: While for college students references are an important non-content feature when assessing credibility, they are not judged as important by high-school students. Hence, it can be assumed that people differing in the ability to evaluate online information differently assess credibility (Lucassen & Schraagen, 2011). Second, it is relevant to know if the identified features are probably an indicator of high or low quality Web sites. Without such knowledge, a successful discrimination between credible and non-credible information can be quite difficult. However, research has proven that users' ability to critically evaluate online information sources can be improved by adequate interventions (e.g., Argelagós & Pifarré, 2012; Braasch, Bråten, Strømsø, Anmarkrud, & Ferguson, 2013; Gerjets & Hellenthal-Schorr, 2008; Walraven, Brand-Gruwel, & Boshuizen, 2008, Walraven, Brand-Gruwel, & Boshuizen, 2010). In order to individually foster a user's ability to evaluate online information and develop adequate training programs, a skill decomposition as well as an adequate assessment tool is needed. Therefore, this study introduces a newly developed assessment tool and explores important conditional cognitive skills needed for the evaluation of online information.

2.4 Definition of Evaluating Online Information

We define the construct *Evaluation of Online Information* as the cognitive ability to efficiently make judgments about the credibility of online information accessed by a search engine by taking structural and message features into account. A competent Web searcher should bear structural and message features in mind while evaluating results on a SERP and the information on corresponding Web sites. Structural features are assumed to be used as predictors for message features both on a SERP and on a selected Web site.

2.5 Cognitive Skills and Abilities Needed for Evaluating Online Search Results

Brand-Gruwel, Wopereis and Walraven (2009) comprehend the critical evaluation of online information as an essential conditional skill in information problem solving (IPS) on the internet. Their IPS-I-model describes this process of information problem solving (IPS) in which the internet (I) is used to search online information as a complex cognitive skill that can be decomposed in main constituent skills, regulation skills and conditional skills. As the three most important conditional skills they name (1) reading skills, (2) evaluating skills and (3) computer skills. In this paper we aim at exploring cognitive skills and abilities that are assumed to contribute to the construct of Evaluating Online Information. According to the IPS-I-model we concentrate on the following conditional skills: basic computer skills because they are specific to the technology of the medium, word processing because it is specific to the modality of the information representation, which is lingual, and at last reasoning ability because it is specific to the way in which the information is processed.

Basic computer skills, defined by Goldhammer, Naumann and Keßel (2013) as "the fundamental ability and speed of performing basic actions in graphical user interfaces of computers to access, collect, and provide information" (p. 264) are assumed to underlie and predict the construct of Evaluating Online Information because they enable to access the Information and Communication Technology (ICT) environment. Involving, amongst others, navigation skills, they are highly beneficial when searching for information online (Brand-Gruwel et al., 2009; Marchionini, 1995; Sutcliffe & Ennis, 1998) and it has been shown that a person's computer experience does matter when evaluating information online (Tu, Shih, & Tsai, 2008). Because online information in contrast to other information sources is typically characterized by a hypertext structure, it is important to have a concept of the hypertext structure and navigation commands in an online environment (Hahnel, Goldhammer, Naumann, & Kröhne, 2016; Goldhammer et al., 2013; International ICT Literacy Panel, 2002). Basic computer skills can be assumed to contribute to ESMF, ESF and EMF because they all require navigation in an online environment when accessing information. However, in particular ESMF is expected to benefit from a high level of basic computer skills because it requires the ability to navigate not only within a Web site, but also between Web sites, i.e., SERPS and linked Web sites which are organized hierarchically. The user needs to know how to use links in a Web environment as well as navigation buttons of a Web browser to efficiently switch back and forth between SERPS and linked Web Sites. In contrast, ESF mostly is required when evaluating information presented on SERPs, and EMF is mainly involved when evaluating online information that is presented in the form of print information as linear text, e.g. when reading newspaper articles or books online. As ESF and EMF refer to the processing of information features of a particular Web site or Web sites at the same level in a hypertext structure (e.g., SERPs), basic computer skills are not expected to contribute as highly to ESF and EMF as to ESMF.

When investigating Web search processes that refer to the cognitive processing of structural and message features, reading literacy, in particular visual word recognition can be assumed to represent a highly relevant cognitive covariate (Brand-Gruwel et al., 2009; Hahnel et al., 2016; Salmerón, Naumann, García, & Farjado, 2017; Kintsch, 1998; Van Dijk & Kintsch, 1983). With the Web being a central aspect of daily life, it is substantially important to be able to read and comprehend information on the Web at high levels (International Reading Association, 2009). But reading online is not the same as reading offline (Coiro, Knobel, Lankshear, & Leu, 2008; International Reading Association, 2002, 2009) because it requires additional skills. According to Castek, Zawilinski, McVerry, O'Byrne, and Leu (2011) reading online requires five processing practices, one of them reading to critically evaluate online information. This practice can be assumed to involve traditional reading skills as phonological encoding, orthographic comparison or the retrieval of word meanings from long-term memory as well as higher-order facets of reading comprehension, such as the semantic and syntactic integration of sentences and the construction of local and global coherence on the text level. Additionally to these reading and comprehension skills that also apply to offline reading, reading to critically evaluate online information can be assumed to involve the judgment of message and structural features. When evaluating the information on SERPs, structural features (e.g. domain names, suffixes) are prominent features. These features present in each link on a SERP need to be scanned and compared, which involves the process of verbal decoding (word recognition). Reading SERPs in a global manner (scanning) before reading them in-depth obviously is important when evaluating online search results (Brand-Gruwel et al., 2009). Therefore it can be assumed that the ability to recognize words and meaningful letter sequences especially predicts ESF. In contrast ESMF and EMF are assumed to involve also higher-order facets of reading literacy, such as the semantic and syntactic integration of sentences and the construction of local and global coherence on the text level, which take place on a higher level of cognitive processing.

Furthermore, reasoning ability is assumed to predict the construct of Evaluating Online Search Results. Research has shown that the ability to critically evaluate information involves amongst others processes of problem-solving (Coiro & et al., 2007). One of the most important cognitive components predicting the competence to solve problems is reasoning ability (Funke & Frensch, 2007). Inductive inferences requiring reasoning ability are ubiquitous and part of our everyday life as most informal arguments are based on them (Johnson-Laird, 1994). When evaluating online search results, making inferences leading to probabilistic conclusions is of great importance, because the mere amount of available information online requires the differentiation between information worth taking a closer look and information that can be ignored. Especially ESF is expected to rely on reasoning ability because it includes the inference of message features from structural features. However, in contrast to ESF, ESMF and EMF may include the direct judgment of message features without any logical conclusions and therefore are not expected to rely as strongly on reasoning ability as ESF. Furthermore, as ESMF is assumed to be predicted by ESF and EMF and the latter in turn are assumed to be predicted by basic computer skills, word recognition and reasoning ability, it can be supposed that the influence of these conditional skills and abilities on ESMF is partly mediated by their effect on ESF and EMF.

2.6 Hypotheses

The first hypothesis addresses the constituent skills of evaluating online information:

(1) ESF and EMF are assumed to represent specific evaluation skills of ESMF. They are assumed to contribute uniquely to ESMF.

The following hypotheses refer to the underlying conditional cognitive skills of ESMF, ESF and EMF:

(2) Basic computer skills are assumed to explain the evaluation of online information. We assume that basic computer skills that is having a notion of a hypertext structure and being able to navigate in an online environment to access information etc., have a stronger direct effect on ESMF than on ESF and EMF due to varying hypertext navigation demands. The direct effect of basic computer skills on ESMF is also assumed to be stronger than possible indirect effects of basic computer skills on ESMF via ESF or EMF.

(3) As ESF includes detecting and reading structural features, e.g. domain names and suffixes, we assume that there is a stronger direct effect of word recognition on ESF than on ESMF and EMF. The indirect effect of word recognition via ESF on ESMF is assumed to be stronger than the indirect effect via EMF on ESMF.

(4) Reasoning ability is assumed to predict ESF because ESF involves the prediction of message features from structural features based on heuristic rules. We assume that the direct effect of reasoning ability on ESMF and EMF is not as strong as the direct effect on ESF because both allow a direct judgment of message features and do not necessarily require making inferences. The indirect effect of reasoning ability via ESF on ESMF is assumed to be stronger than the indirect effect via EMF on ESMF. Figure 1 presents the research model.

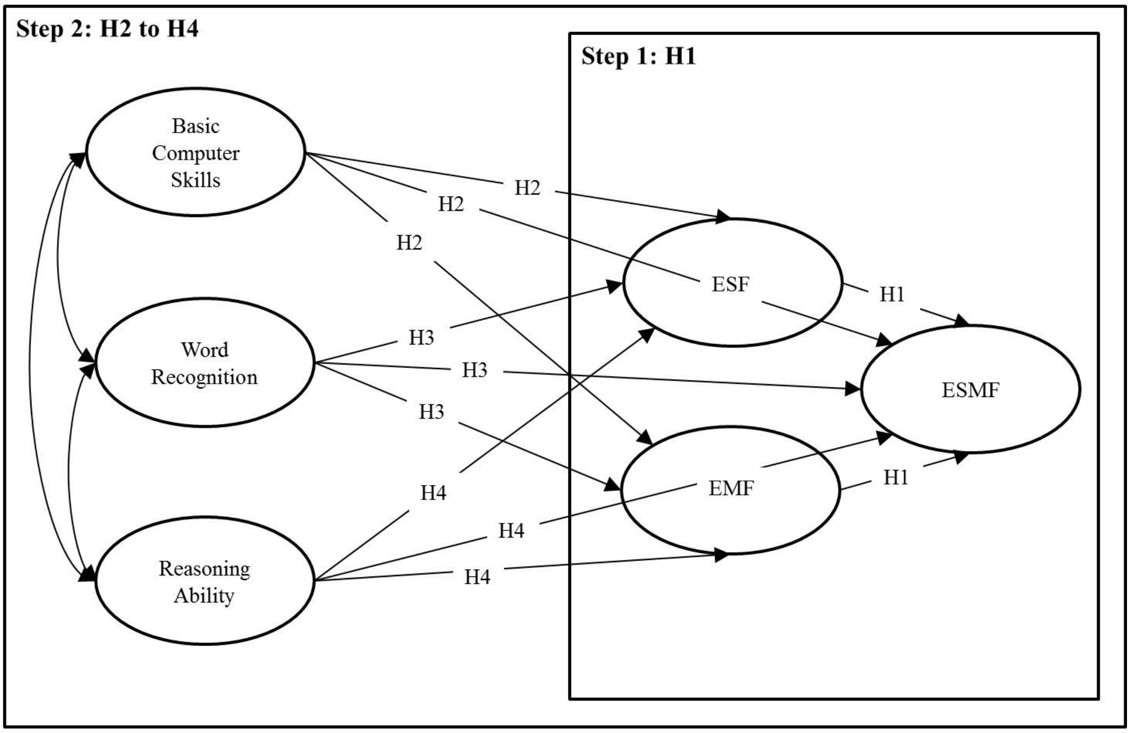


Figure 1. Research model (latent level).

3. Material and methods

3.1 Participants

Participants were 205 high school students in grade 12 (42.4% female, 57.6% male) at nine schools in Germany, with an overall mean age of 18.03 (*SD* = .70) ranged from 15.5 to 21.75 years.

3.2 Measured Variables

3.2.1 Test for evaluating online information (TEO).

The cognitive ability to efficiently make judgments about the credibility of online information using structural and message features was assessed by the Test for Evaluating Online Information (TEO), which was newly developed for this study. The test consists of 24 interactive items simulating Web environments. The 24 items are assigned to three subscales ESMF, ESF and EMF, with eight items per scale (cf. Table 1).

Table 1. TEO Test Development.

			Dimensions of Task Design			Context Factors	
Subscale	Item	Included Credibility Features in Online Sources	Attractiveness of Non-Targets	Congruency of Structural and Message Features	Number of Online Sources to be evaluated	Text form	Topic
ESF	1	Structural Features	Low	_	6	SERP	Health
ESF	2	Structural Features	Low	_	10	SERP	Crafts
ESF	3	Structural Features	High	_	10	SERP	Sports
ESF	4	Structural Features	High	_	6	SERP	Education
ESF	5	Structural Features	Low	_	6	SERP	Crafts
ESF	6	Structural Features	Low	_	10	SERP	Sports
ESF	7	Structural Features	High	_	10	SERP	Education
ESF	8	Structural Features	High	_	6	SERP	Health
ESMF	9	Structural + Message Features	High	Yes	5	SERP + Web sites	Sports
ESMF	10	Structural + Message Features	High	Yes	3	SERP + Web sites	Education
ESMF	11	Structural + Message Features	High	No	5	SERP + Web sites	Health
ESMF	12	Structural + Message Features	High	No	3	SERP + Web sites	Crafts
ESMF	13	Structural + Message Features	Low	Yes	5	SERP + Web sites	Education

ESMF	14	Structural + Message Features	Low	Yes	3	SERP + Web sites	Health
ESMF	15	Structural + Message Features	Low	No	5	SERP + Web sites	Crafts
ESMF	16	Structural + Message Features	Low	No	3	SERP + Web sites	Sports
EMF	17	Message Features	Low	-	5	Internet forum	Health
EMF	18	Message Features	Low	-	3	Internet forum	Crafts
EMF	19	Message Features	High	-	5	Internet forum	Sports
EMF	20	Message Features	High	-	3	Internet forum	Education
EMF	21	Message Features	Low	-	5	Internet forum	Education
EMF	22	Message Features	Low	-	3	Internet forum	Sports
EMF	23	Message Features	High	-	3	Internet forum	Education
EMF	24	Message Features	High	-	5	Internet forum	Health

The subscale ESMF consists of eight items with online information presented in the form of Google-like SERPs and corresponding Web sites. The items contain either three or five links on one SERP and a corresponding Web site for each of the links. Structural and message features are available for evaluation. Participants first see the SERP with a search term already given and three or five links. By clicking on these links the corresponding Web sites can be reached. These Web sites are static sites with all hyperlinks, buttons or menus being inactive. The online information included in the items consists of factual texts, which can be considered either to be true or false. Participants are required to identify the most credible link/ Web site by selecting a radio button next to the link of their choice. For each task the individual response is collected. There is only one correct response. The correct response is defined by the link/ Web site with the highest number of information features indicating credibility. Selecting the correct link/ Web site is scored with 1, selecting a false link/ Web site is scored with 0.

The eight items of the subscale ESF consist of either six links on one SERP or ten links on two linked SERPS without corresponding Web sites, i.e., the test person cannot not reach a corresponding Web site by clicking on a search result. In this case the evaluation of search results is based on sparse information about the corresponding Web sites and the information they contain. Mostly structural features are available for the evaluation. Again, participants are required to identify the most credible link by selecting a radio button next to their link of choice on the SERP. For each task the individual response is collected. There is only one correct response defined by the link with the highest number of information features indicating credibility. Selecting the correct link is scored with 1, selecting a false link is scored with 0.

A typical SERP for the subscales ESMF and ESF is displayed in Figure 2. The correct answer for the sample item presented in Figure 2 is the third link.

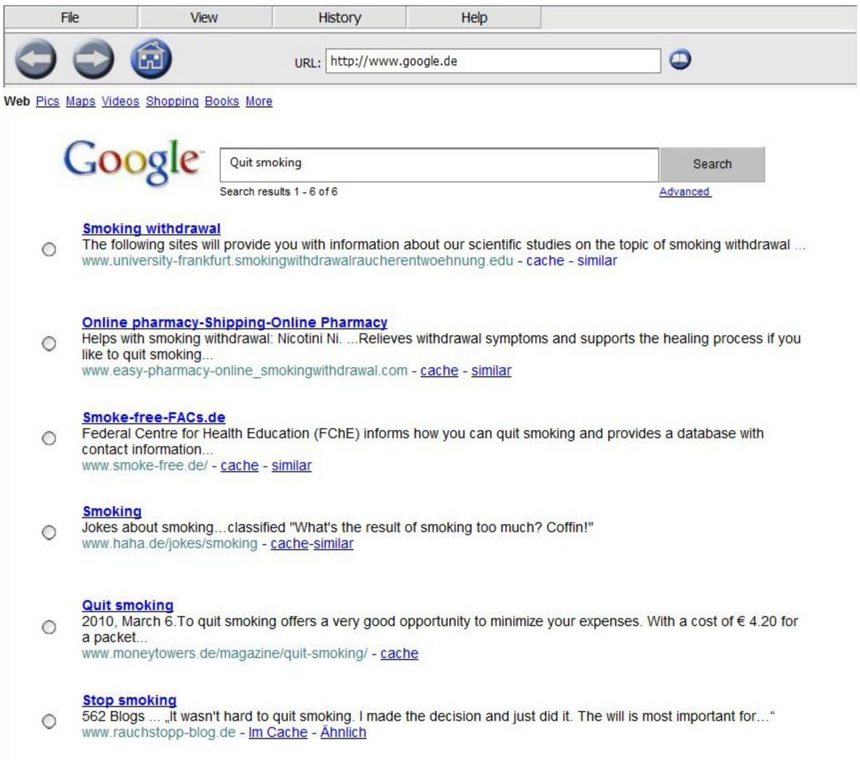


Figure 2. Screenshot of a SERP for the topic “smoking cessation” with six search results.

The EMF subscale comprises eight items. Each item represents an internet forum with three or five forum entries answering a question related to the topics health, crafts, sports or education. Participants are asked to identify the most credible entry by selecting a radio button next to it. For each task the individual response is collected. There is only one correct response defined by the forum entry with the highest number of information features indicating credibility. Selecting the correct response is scored with 1, selecting a false response is scored with 0. A typical item for the EMF subscale is displayed in Figure 3. The correct answer for this item is the third forum entry.

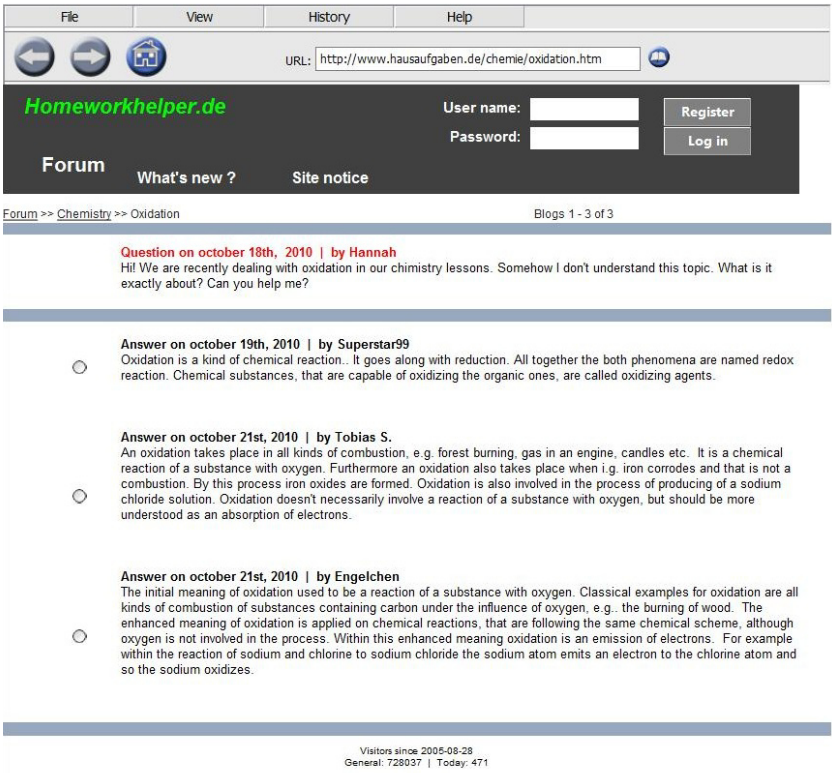


Figure 3. Screenshot of an internet forum with three entries relating to the topic "oxidation".

For all items the positions of the correct and false responses on the SERP and on the forum site were random. The items of the three subscales were designed by manipulating task characteristics that were assumed to affect item difficulty. The following attributes are balanced across the items as shown in Table 1: the attractiveness of non-target search results or forum entries (high vs. low), the congruency of structural and message features, which is only applicable to the items of the ESMF subscale (yes vs. no) and the number of information sources to be compared (three vs. five for the ESMF scale, six vs. ten for the ESF and EMF scales).

If the attractiveness of the non-target search results or non-target forum entries compared to the target search result or target forum entry is high, they only differ in very few structural and or message features indicating differences in credibility, and therefore they are more difficult to distinguish (and vice versa). In Figure 2, the attractiveness of the non-targets of this sample SERP is low. In Figure 3, the attractiveness of the non-targets of these sample forum entries item is high.

The congruency of structural and message features is varied for the ESMF scale by designing one group of (congruent) items with structural and message features both indicating credibility or both indicating non-credibility, and another group of (incongruent) items with structural features indicating credibility and message features indicating non-credibility. Congruency of structural and message features (yes vs. no) was verified based on expert ratings. For incongruent items the correct response was defined according to the features on the Web site, where structural and message features are present.

Concerning the number of information sources an item is assumed to be more difficult, if multiple information sources need to be compared. The search topics were also balanced across items. Search topics are health, crafts, sports and education. They were selected to minimize the influence of background knowledge. Search topics were not expected to have an impact on item difficulty.

3.2.2 Basic computer skills.

Basic computer skills were assessed with a revised version of the scale *Basic Computer Skills* (BCS) by Goldhammer et al. (2013) comprising 20 interactive items assessing the ability to operate mouse and keyboard, to open, save, delete and change texts and to handle menu bars, hyperlinks, scrollbars and hotkeys. Items include simulated computer environments representing the standard programs, i.e., word processing, Web browser and email client. Reliability analysis revealed a Cronbach's alpha of .69 for the Basic computer skills scale. Items fit a unidimensional 2 Parameter Logistic (2PL)-model.

3.2.3 Word recognition.

Student's accuracy of word recognition was assessed with a Lexical Decision Task (Balota, Cortese, Sergent-Marshall, & Yap, 2004; Richter, Naumann, Isberner, & Kutzner, 2011). Students judged eight words and eight non-words (16 items) that appeared successively on a computer screen. Participants were asked to work as quickly as possible, while avoiding errors. Individual differences in word recognition ability are reflected by correct and incorrect responses. Each of the eight words were nouns. They varied in length between three and ten letters and between one and three syllables. To cover a wide variety of difficulty levels, also orthographic regularity and number of orthographic neighbours were varied across items (Andrews, 1989; Balota et al., 2004; Huntsman & Lima, 2002). Non-words were constructed by finding for each word another word that was equivalent in word type, length, frequency, regularity and number of orthographic neighbours. From this second word then a non-word was constructed by switching letters. Reliability analysis revealed a Cronbach's alpha of .74 for the Word Recognition Test. Items fit a unidimensional 2PL-model.

3.2.4 Reasoning ability.

Reasoning ability was assessed with 16 items measuring fluid intelligence from the Berlin Test of Fluid and Crystallized Intelligence (BEFKI) (Schipolowski, Schroeders, & Wilhelm, 2008). Participants have to complete rows of geometric objects based on rules to be induced. The items consist of rows with either two, three or four geometric objects with two empty positions in the end. The test taker is required to complete each empty position by choosing among three possible alternative geometric objects. Reliability analysis for the 16 items revealed Cronbach's alpha to be .67. The 16 items of the BEFKI have shown to fit a unidimensional 1PL-model.

3.3 Procedure

Data was collected in the context of a pilot study of the *National Educational Panel Study* (NEPS) sponsored by the Federal Ministry of Education and Research in Germany. Tests were administered in groups of up to 24 students in computer labs at schools. The TEO was administered at the beginning of the test session. On average students completed the 24 items of the TEO in 18 minutes.

Besides the mentioned tests students also completed two other tests: a modified version of the Frankfurt Adaptive Concentration Test (Moosbrugger & Goldhammer, 2007) and the Pracowi test measuring practical computer knowledge (Naumann, Richter, & Groeben, 2001), however these tests were not included in the present study. The overall completion time was about 120 minutes with a 5 minute break after 60 minutes.

3.4 Data Analysis

Analyses are based on both dichotomous item responses and item parcels. Following Little, Cunningham, Shahar and Widaman (2002) item responses are modeled to investigate the structure of items, whereas item parcels are modeled to investigate the structure of constructs.

Psychometric properties of the TEO subscales and items' properties were analysed using Confirmatory Factory Analysis (CFA) for categorical data, i.e., dichotomous items (Bollen, 1989), and, by computing the item coefficients difficulty and discrimination as proposed by classical test theory (CTT). CFA was used instead of explanatory factor analysis (EFA), because hypotheses about the latent factor structure and the relation of items to factors were a priori available, i.e., they were suggested by the theoretical rationale underlying the development of TEO.

Structural equation models (SEM) with item parcels as indicators were used to investigate the association of basic computer skills, word recognition and reasoning ability with ESMF, ESF and EMF. Item parcels were constructed to obtain normally distributed indicator variables, fewer and more stable parameter estimates and reduced sources of sampling error (Holt, 2004; Little et al., 2002; Little, Rhemtulla, Gibson, & Schoemann, 2013). For each test three item parcels were created. As recommended by Little et al. (2002) first the factor loadings of the items of a scale were investigated. Then items with high factor loadings, items with moderate factor loadings and items with low factor loadings were grouped together. After that items from each of the three groups were randomly assigned to three different item parcels.

To conduct CFAs including categorical item response indicators, the WLSMV estimator was selected (Forero & Maydeu-Olivares, 2009). For CFAs and SEMs including cognitive covariates and continuous indicators such as item parcels, the standard ML estimator was used. A model was considered to show a good/acceptable overall model fit if the model fit criteria according to Muthén (1998-2004) and Schermelleh-Engel, Moosbrugger & Müller (2003) were met. To test if two nested models differ significantly, the Wald Test was used (Bollen, 1989). For the statistical tests of significance an alpha level of .05 was assumed. Analyses taking the clustering of students into schools into account were not conducted due to the small number of clusters which would result in unreliable standard error corrections. Model parameters were estimated by means of the Mplus software version 6 (Muthén & Muthén, 1998-2011). Item coefficients related to CTT were estimated using the ltm package (Rizopoulos, 2006) of the R environment (R Development Core Team, 2011).

4. Results

For the three TEO subscales, CFA model parameters as well as CTT item coefficients can be found in the appendix. Furthermore, the appendix comprehends descriptive statistics for the variables ESMF, ESF, EMF, basic computer skills, word recognition and reasoning. Results for the initially tested CFA model showed that the items 10 and 16 of the ESMF scale did not load significantly on the ESMF factor and showed low discrimination coefficients. Therefore they were dropped. Item 8 of the ESF scale reduced the model fit significantly and was dropped as well.

Reliability analysis revealed a reliability of .52 (Cronbach's α value) for the subscale ESMF. Factor loadings for this subscale ranged between .19 (n.s.) and .81. Items' proportion of correct responses ranged between .41 and .90, with 4 of the six remaining items showing a proportion of correct responses above .50. That means items were rather easy with most of the participants being able to solve most of the items. The subscale ESF showed a Cronbach's α of .58. Factor loadings ranged between .51 and .84. Items' proportion of correct responses ranged between .41 and .94, with six of the remaining seven items showing a proportion of correct responses above .50. Again, items were rather easy. Results for the EMF subscale revealed a Cronbach's α value of .52. Factor loadings ranged between .18 (n.s.) and .82. Items' proportion of correct responses ranged between .82 and .95, indicating that again items were easy and most of the participants were able to solve most of the items.

4.1 Results on Hypothesis 1

To test Hypothesis 1 a latent regression of ESMF on ESF and EMF was conducted. The model fit the data very well, χ^2 (*df*=186, N=205)= 192.09, *p*= .36, RMSEA= .013, WRMR= .08, CFI=.98, TLI=.98. ESF could be identified as a predictor for ESMF, b=.65, SE(b)=.12, p<.001. EMF was also found to predict ESMF, b=.32, SE(b)=.12, p=.009. In the regression model 76% of the variance of ESMF (R^2=.76, *p*<.001) was explained by the dimensions ESF and EMF. Thus, Hypothesis 1 was supported by the data.

4.2 Contribution of Basic Computer Skills, Reasoning Ability and Word Recognition to ESMF, ESF and EMF

To get an overview of the correlation structure of ESMF, ESF and EMF with basic computer skills, word recognition and reasoning ability at the latent level, a multidimensional CFA-model with all cognitive covariates was tested. The model fit the data very well, χ^2 (df=120, N=205)= 123.22, p= . 40, RMSEA= .011, SRMR= .05, CFI=.99, TLI=.99. As expected and shown in Table 2, ESF showed higher correlations with the covariates word recognition and reasoning ability than ESMF and EMF. The correlations of ESF, ESMF and EMF with basic computer skills did not differ substantially: All correlations of ESF with the covariates were significant. ESMF and EMF did not show a significant correlation with word recognition.

Table 2. Latent correlations of ESMF, ESF and EMF with covariates.

	ESMF	ESF	EMF
Basic Computer Skills	.39*	.41*	.37*
Word Recognition	.19 (n.s.)	.42*	.15 (n.s.)
Reasoning Ability	.32*	.56*	.35*

Note: *p<.05, n.s.= not significant

4.2.1 Results on Hypothesis 2.

To address Hypothesis 2, a latent path analysis regressing ESMF, ESF and EMF on basic computer skills, word recognition and reasoning ability was conducted with the relationship between ESMF and the cognitive covariates being also mediated by ESF and EMF. The mediation model fit the data very well, χ^2 (df=121, N=205)= 129.54, p= .28, RMSEA= .019, SRMR= .05, CFI=.98, TLI=.98. In the mediation model 61% of the variance of ESMF (R^2=.61, p<.001) was explained by ESF, EMF and the cognitive covariates basic computer skills, word recognition and reasoning ability. Furthermore, 50% of the variance of ESF (R^2=.50, p<.010) and 22% of the variance of EMF (R^2=.22, p<.001) was explained by the covariates basic computer skills, word recognition and reasoning ability. In Hypothesis 2 the direct effect of basic computer skills on ESMF was assumed to be stronger than the direct effect of basic computer skills on ESF and EMF as well as possible indirect effects. Results are presented in Table 3. The mediation analysis did not reveal any significant direct or indirect effects of basic computer skills. Thus, Hypothesis 2 was not supported by the results.

Table 3. Results on Hypothesis 2, 3 and 4: Latent Mediation Model.

	Criterion					
Predictors	**ESMF**		**ESF**		**EMF**	
	β	z	β	z	β	z
Basic Computer Skills	.14	.87	.12	.82	.22	1.38
Word Recognition	-.20	-1.38	.34**	3.39	.10	.83
Reasoning Ability	-.31	-1.43	.51***	3.89	.30*	2.03
ESF	.86***	4.13	-	-	-	-
EMF	.23	1.75	-	-	-	-
Basic Computer Skills via EMF	.05	1.14	-	-	-	-
Basic Computer Skills via ESF	.10	.83	-	-	-	-
Word Recognition via EMF	.02	.74	-	-	-	-
Word Recognition via ESF	.29*	.2.39	-	-	-	-
Reasoning Ability via EMF	.07	1.26	-	-	-	-
Reasoning Ability via ESF	.44*	2.38	-	-	-	-

4.2.2 Results on Hypothesis 3.

In Hypothesis 3 it was assumed that there is a stronger direct effect for word recognition on ESF than on ESMF and EMF, because ESF mainly includes detecting and decoding verbal structural features, while ESMF and EMF also require higher-level text processing. Furthermore, the assumed indirect effect of word recognition on ESMF via ESF was expected to be stronger than the assumed indirect effect of word recognition on ESMF via EMF. Results can be found in Table 3. In line with Hypothesis 3, the latent mediation analysis revealed a significant direct effect for word recognition on ESF. There was no significant direct effect for word recognition on ESMF and EMF. The model-difference-test revealed the difference in the direct effects of word recognition on ESF and ESMF to be significant (Wald Test: W= 4.58, df=1, p<. 05). The difference in the direct effects for word recognition on ESF and EMF was not significant (Wald Test: W= 3.67, df=1, p=.06). As expected, the path analysis revealed a significant indirect effect of word recognition on ESMF via ESF. The indirect effect of word recognition on ESMF via EMF was not significant (cf. Table 3). Thus, Hypothesis 3 was generally supported by the results.

4.2.3 Results on Hypothesis 4.

In Hypothesis 4 it was claimed that the direct effect of reasoning ability on ESF is stronger than the direct effect of reasoning ability on ESMF and EMF because ESF involves the application of heuristic rules to make inferences, which ESMF and EMF do not (necessarily) include. In addition the assumed indirect effect of reasoning ability via ESF on ESMF was expected to be stronger than the assumed indirect effect via EMF on ESMF. Supporting Hypothesis 4, the latent path analysis showed a significant direct effect of reasoning ability on ESF. A significant direct effect of reasoning ability was also revealed on EMF, whereas there was no significant direct effect of reasoning ability on ESMF (cf. Table 3). A model-difference-test uncovered that the difference in the effects of reasoning ability on ESF and on ESMF was significant (Wald Test: W= 4.47, df=1, p<.05). The difference in the effects of reasoning ability on ESF and EMF did not prove to be significant though (Wald Test: W= 2.16, df=1, p=.14). As expected, the latent mediation analysis revealed a significant indirect effect of reasoning ability on ESMF via ESF, while the indirect effect of reasoning ability on ESMF via EMF was not significant. That is, Hypothesis 4 could generally be supported by the results.

5. Discussion and Conclusions

The overall goal of the present study was to present a newly developed scale measuring the ability of evaluating online information and to further clarify the nature of the construct both theoretically and empirically. We defined evaluating online information as the cognitive ability to efficiently evaluate the credibility of online information including the judgment of message and structural features. Based on that, we applied our newly developed measurement tool to explore the latent structure of the construct Evaluating Online Information, requiring the evaluation of both structural and message features (ESMF), and its constituent skills, that is, ESF, requiring the evaluation of structural features only and EMF, requiring the evaluation of message features only. Furthermore, we investigated the relation of the cognitive covariates basic computer skills, word recognition and reasoning ability with ESMF, ESF and EMF.

In line with Hypothesis 1 ESF and EMF revealed to be constituent skills of ESMF, indicating that the evaluation of online information underlies different skills based on which features of online information are paid attention to and are evaluated with respect to their credibility. Given the results from path analysis, ESMF, supposed to incorporate both ESF and EMF, more strongly relies on ESF than on EMF. This may be due to a hierarchy in the evaluation process of online search results, in which structural features are evaluated first, followed by message features. Although the data used in this study cannot prove such conclusions, future research should explore this posited sequence. This hierarchy of evaluation of course may be limited to the evaluation of online information, in which structural features are the salient information characteristics on SERPs, which are accessed first during the Web search process, and, therefore scrutinized more strongly. Though, it could be possible that other types of online information show a different hierarchy of evaluation.

Results related to Hypotheses 2 to 4 provide insights into cognitive skills underlying the evaluation of online search results. In contrast to our assumptions made in Hypothesis 2, basic computer skills did not contribute uniquely to ESMF and its constituents ESF and EMF. In consequence, instructions and training programs on how to evaluate online information do not necessarily have to incorporate training in computer skills, as they do not seem to be an essential component and especially younger students nowadays may already possess sufficient computer experience (Aslanidou & Menexes, 2008).

Hypothesis 3 and 4 were supported by the data. As supposed, word recognition has a strong effect on ESF. We explain this strong effect of word recognition by the process underlying the evaluation of structural features. They need to be scanned and compared, which involves verbal decoding and the recognition of words and meaningful letter sequences as domain names or suffixes. These findings are in line with the finding of Brand-Gruwel et al. (2009), who have shown that reading globally (scanning) occurs much more often than reading in-depth while evaluating information on the internet and is very important in this process. Furthermore, reasoning ability was found to have a strong effect on ESF as well. Our interpretation for this effect is that structural features are used to infer message features based on heuristic rules when evaluating information. Reasoning ability is also a significant predictor for EMF because message features need to be analysed with regard to their quality, although as expected the effect is smaller than the effect on ESF. There was no direct effect of reasoning ability on ESMF. To ensure that this might not be due to the design of the ESMF subscale, as half of the items contained incongruent structural and message features, which makes drawing inferences impossible, we re-analysed the data without the four incongruent items, but the effect of reasoning ability on ESMF remained insignificant. However, while word recognition and reasoning ability do not directly predict ESMF, they have significant indirect effects on ESMF via ESF.

Word recognition was assumed to be involved in the evaluation of online information with regard to the processing of structural features. This could be confirmed empirically. As a next step, for future research it would be rewarding to investigate the relation of ESMF, ESF and EMF with reading comprehension, of which word recognition is an important component. Hahnel et al. (2016) have shown, that competent readers select and visit more websites containing relevant information. Thus, higher-order facets of reading comprehension may have an impact on the three dimensions as well and predict the successful completion of the TEO items incrementally. Poor readers can be assumed to concentrate more on text comprehension and therefore have less capacity for the evaluation processes during Web search than fluent readers. In consequence fluent readers might have an advantage over poor readers when evaluating online search results. Research in the field of reading comprehension has shown repeatedly that poor readers tend to focus on the decoding of individual words at the expense of deeper comprehension (Perfetti, 1985). Therefore, there might be no difference in evaluating structural features between poor and fluent readers, but the evaluation of message features may pose a greater challenge to poor readers than to fluent readers. Because ESMF and EMF mainly involve high-order facets of reading literacy (reading in-depth) in contrast to ESF, reading comprehension is expected have a stronger effect on ESMF and EMF than on ESF.

A limitation of this study and also a possible explanation that regarding Hypothesis 3 and 4 some differences in effects were not revealed to be significant, is the homogeneity of the sample (only grade 12high-school students), which also limits the generalizability of the results. Correlations and regression effects might be stronger when using a more heterogeneous sample, which is suggested for future research. In addition, as the school system in Germany differs from school systems in other countries, results are restricted to German high-school students. Another, related limitation is sample size of the present study. Although, it can be considered to be typical for SEM studies (Kline, 2011), it should be increased to obtain more accurate and stable parameter estimates in complex models. The reliabilities (Cronbach's α) of the three subscales turned out to be rather low, which might be due to the small number of items included in each subscale. Therefore, in future research aiming at the accurate assessment of individuals, the subscales should be extended by developing new items.

Furthermore, in the recent past the structure of SERPS has undergone some changes in that they increasingly include images and videos in addition to text. For sure, this may have an impact on the process of link selection and credibility assessment, and, therefore, new items should include this extra information on SERPs. Moreover, the present study can neither claim to be exhaustive with respect to the investigated credibility features nor show directly their relevance for the evaluation of online search results. Therefore, future research using the TEO should incorporate thinking aloud-protocols and eye-tracking data to gain deeper insights of the actual cognitive processes carried out (Cutrell & Guan, 2007, Kammerer & Gerjets, 2012; Van Gog, Paas, & van Merriënboer, 2005). In addition, students' background knowledge about the item topics should be assessed and explicitly controlled for.

Findings of the present study contribute to the clarification of the theoretical background of the ability of evaluating online information. It could be shown that the evaluation of online information (ESMF) is a complex construct underlying different evaluation skills (ESF and EMF) as well as the cognitive skills of word recognition and reasoning ability.

Research has shown that the amount of experience with the Web does not have an impact on how well people evaluate the credibility of online information (Fogg et al., 2001) and that students, who are usually experienced using the Web, appear to use only very few criteria when evaluating online information (Brand-Gruwel et al., 2009). One of these few criteria is usability of information that is very often evaluated, while reliability is not (Brand-Gruwel et al., 2009). A successful information evaluation on the Web could therefore be fostered by teaching students which criteria (structural and message features of SERPs and Web sites) they need to pay attention to and how to rate these features. As shown empirically, especially word recognition and reasoning ability can be considered prerequisite skills for Web-based information evaluation. From an instructional point of view, these findings may foster the development of adequate interventions, which is imperative because research has repeatedly shown that most people lack the skills to appropriately evaluate online information (Brand-Gruwel et al., 2005; Gerjets & Hellenthal-Schorr, 2008; Rouet, Ros, Goumi, Macedo-Rouet, & Dinet, 2011). Trainings for Web users might benefit from including the improvement of the basic cognitive skills underlying the ability of information evaluation. This particularly refers to cognitive skills that can be modified by learning and instruction such as basic reading skills, and the knowledge about and the use of structural features.

For further research it would be interesting to study individual differences in strategies of evaluating online information. Process variables derived from logfiles could shed some light on the differences between more and less competent online information evaluators. Variables such as the time spent looking at SERPs and Web sites, and the number of visited Web sites might determine a successful information evaluation (OECD, 2011). Such analyses could possibly identify effective evaluation strategies that can be incorporated in Web search trainings for students and adults. Moreover, process variables could be used to further distinguish evaluators showing the same ability level with regard to their efficiency.

The presented findings enable new insights into the complex construct of evaluating online search results. Moreover, we expect that these findings can bring forward adequate instructions to foster important conditional skills making information search on the Web a benefit for everyone rather than a challenge.

References

Andrews, S. (1989). Frequency and neighbourhood size effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 802-814.

Argelagós, E., & Pifarré, M. (2012). Improving information problem solving skills in secondary education through embedded instruction. *Computers in Human Behavior, 28*, 515–526.

Aslanidou, S., & Menexes, G. (2008). Youth and the internet: Uses and practices in the home. *Computers and Education, 51*, 1375–1391.

Balota, D.A., Cortese, M.J., Sergent-Marshall, S.D. & Yap, M.J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133*, 283-316.

Bollen, K.A. (1989). *Structural equations with latent variables.* New York: Wiley.

Braasch, J. L. G., Bråten, I., Strømsø, H. I., Anmarkrud, Ø., & Ferguson, L. E. (2013). Promoting secondary school students' evaluation of source features of multiple documents. *Contemporary Educational Psychology, 38*, 180–195.

Brand-Gruwel, S., Wopereis, I., & Vermetten, Y. (2005). Information problem solving by experts and novices: analysis of a complex cognitive skill. *Computers in Human Behavior, 21*, 487-508.

Brand-Gruwel, S., Wopereis, I., & Walraven, A. (2009). A descriptive model of Information Problem Solving while using Internet. *Computers & Education, 53*, 1207-1217.

Brand-Gruwel, S., & Stadtler, M. (2011). Solving information-based problems: Searching, selecting and evaluating information. *Learning and Instruction, 21*, 175 –179

Brand-Gruwel, S., Kammerer, Y., van Meeuwen, L., & van Gog, T. (2017). Source evaluation of domain experts and novices during Web search. *Journal of Computer Assisted Learning, 33*, 234-251.

Britt, M.A., & Aglinskas, C. (2002). Improving student's ability to identify and use source information. *Cognition & Instruction, 20*, 485-522.

Burkell, J. (2004). Health information seals of approval: What do they signify? Information, *Communication, & Society, 7*, 491–509.

Castek, J., Zawilinski, L., McVerry, G., O'Byrne, I., & Leu, D. (2011). The new literacies of online reading comprehension: New opportunities and challenges for students with learning difficulties. In: C. Wyatt-Smith, J. Elkins, & S. Gunn (Eds.) *Multiple Perspectives on Difficulties in Learning Literacy and Numeracy* (pp. 91-110). New York, NY: Springer.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology, 39*(5), 752-766.

Coiro, J., & Dobler, E. (2007). Exploring the online reading comprehension strategies used by sixth-grade skilled readers to search for and locate information on the internet. *Reading Research Quarterly, 42*, 214-257.

Coiro, J., Knobel, M., Lankshear, C. &, Leu, D. (2008). Central issues in new literacies and new literacies research. IN: J. Coiro, M. Knobel, C. Lankshear and D. Leu (Eds.) *Handbook of research on new literacies* (pp. 1-31). New York: Lawrence Erlbaum Associates.

Cutrell, E. & Guan, Z. (2007). What are you looking for? An eye-tracking study of information usage in Web search. *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, 407-416.

Fogg, B.J., & Tseng, S. (1999). The elements of computer credibility. *Proceedings of Human Interface SIG conference*, 80-87.

Fogg, B.J., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., Paul, J., Rangnekar, A., Shon, J., Swani, P., & Treinen, M. (2001). What Makes A Web Site Credible? A Report on a Large Quantitative Study. *Proceedings of ACM CHI 2001 Conference on Human Factors in Computing Systems* (61-68). New York: ACM Press. Available at http://www.acm.org/pubs/articles/proceedings/chi/365024/p61-fogg/p61-fogg.pdf.

Fogg, B.J. (2002). Prominence-Interpretation Theory: Explaining How People Assess Credibility. A Research Report by the Stanford Persuasive Technology Lab, Stanford University. Available at http://credibility.stanford.edu/pit.html or http://www.webcredibility.org/pit.html.

Forero, C.G. & Maydeu-Olivares, A. (2009). Estimation of IRT graded models for rating data: Limited vs. full information methods. *Psychological Methods, 14*, 275-299.

Funke, J. & Frensch, P.A. (2007). Complex problem solving: The European perspective - 10 years after. In: David H. Jonassen (Ed.) *Learning to solve complex scientific problems* (pp. 25-47). New York: Lawrence Erlbaum Associates. Available at: http://works.bepress.com/joachim_funke/18

Gerjets, P., & Hellenthal-Schorr, T. (2008). Competent information search in the World Wide Web: development and evaluation of a Web training for pupils. *Computers in Human Behavior, 24*, 693-715.

Gerjets, P., Kammerer, Y., & Werner, B. (2011). Measuring spontaneous and instructed evaluation processes during Web search: Integrating concurrent thinking-aloud protocols and eye-tracking data. *Learning and Instruction, 21*, 220-231.

Gigerenzer, G. & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103*(4), 650-669.

Go, E., You, K.H., Jung, E., & Shim, H. (2016). Why do we use different types of websites and assign them different levels of credibility? Structural relations among users' motives, types of websites, information credibility, and trust in the press. *Computers in Human Behavior, 54*, 231-239.

Goldhammer, F., Naumann, J. & Keßel, Y. (2013). Assessing Individual differences in Basic Computer Skills: Psychometric characteristics of an interactive performance measure. *European Journal of Psychological Assessment, 29*, 263-275.

Hahnel, C., Goldhammer, F., Naumann, J., & Kröhne, U. (2016). Effects of linear reading, basic computer skills, evaluating online information, and navigation on reading digital text. *Computers in Human Behavior, 55*, 486-500.

Hargittai, E., Fullerton, L., Menchen-Trevino, E., & Thomas, K. Y. (2010). Trust Online: Young Adults' Evaluation of Web Content. *International Journal of Communication, 4*, 468-494.

Holt, J. K. 2004. *Item Parceling in Structural Equation Models for Optimum Solutions.* Paper presented at the 2004 Annual Meeting of the Mid-Western Educational Research Association, Columbus, OH. (Paper available upon request.)

Hong, T. (2006). The influence of structural and message features on website credibility. *Journal of the American Society for Information Science and Technology, 57*(1), 114-127.

Huntsman, L.A. & Lima, S.D. (2002). Orthographic neighbours and visual word recognition. *Journal of Psycholinguistic Research, 31*, 289-306.

International ICT Literacy Panel (2002). Digital transformation: A framework for ICT literacy (A report of the International ICT Literacy Panel). Princeton, NJ: Educational Testing Service. Available from: http://www.ets.org/Media/Tests/Information_and_Communication_Technology_Literacy/ictreport.pdf [Accessed 17 October 2012].

International Reading Association (2002). *Integrating literacy and technology in the curriculum: A position statement.* Newark, DE: International Reading Association.

International Reading Association (2009). *IRA Position Statement on New Literacies: A Position Statement.* Newark, DE: International Reading Association.

Johnson-Laird, P.N. (1994). Mental models and probabilistic thinking. *Cognition, 50*, 189-209.

Kammerer, Y., & Gerjets, P. (2012). Effects of search interface and Internet-specific epistemic beliefs on source evaluations during Web search for medical information: an eye-tracking study. *Behaviour & Information Technology, 31*(1), 83-97.

Kintsch, W. (1998) *Comprehension: A paradigm for cognition.* New York: Cambridge University Press.

Kline, R.B. (2011). *Principles and Practice of Structural Equation Modeling.* 3rd ed. New York: Guilford Press.

Leu, D.J., Zawilinski, L., Castek, J., Banerjee, M., Housand, B., Liu, Y., & O'Neil, M. (2007). What is new about the new literacies of online reading comprehension? In: J. Eakle, L. Rush, & A. Berger (Eds.) *Secondary school literacy: What research reveals for classroom practice* (pp. 37-68). Urbana, IL: National Council of Teachers of English (NCTE).

Lewandowski, D. (2011). The influence of commercial intent of search results on their perceived relevance. In S. Sadagopan & K. Ramamritham (Eds.), *Proceedings of the 2011 iConference* (iConference '11; pp. 452–458). NewYork, NY: ACM Press.

Lewandowski, D. (2013). Credibility in Web search engines. In S. Apostel & M. Folk (Eds.), *Online credibility and digital ethos: Evaluating computermediated communication* (pp. 131–146). Hershey, PA: IGI Global

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling, 9*, 151–173.

Little, T. D., Rhemtulla, M., Gibson, K., & Schoemann, A. M. (2013). Why the Items Versus Parcels Controversy Needn't Be One. *Psychological Methods, 18*(3), 285-300.

Lorenzen, M. (2001). The land of confusion? High school students and their use of the World Wide Web for research. *Researching Strategies, 18*, 151-163.

Lucassen, T., & Schraagen, J.M. (2010). Trust in Wikipedia: How users trust information from an unknown source. In *Proceedings of the Fourth Workshop on Information Credibility* (WICOW '10) (pp. 19–26). New York: ACM Press.

Lucassen, T. & Schraagen, J.M. (2011). Factual accuracy and trust in information: The role of expertise. *Journal of the American Society for Information Science and Technology, 62*(7), 1232-1242.

MaKinster, J. G., Beghetto, R. A., & Plucker, J. A. (2002). Why can't I find Newton's third law? Case studies of students' use of the web as a science resource. *Journal of Science Education and Technology, 11*, 155–172.

Marchionini, G. (1995). *Information-seeking in electronic environments.* New York: Cambridge University Press.

Meola, M. (2004). Chucking the checklist: A contextual approach to teaching undergraduates Web-site evaluation. *Libraries and the Academy, 4*(3), 331-344.

Mercier, J. & Frederiksen, C. (2008). The structure of the help-seeking process in collaboratively using a computer coach in problem-based learning. *Computers & Education, 51*, 17-33.

Metzger, M. (2007). Making Sense of Credibility on the Web: Models for Evaluating Online Information and Recommendations for Future Research. *Journal of the American Society for Information Science and Technology, 58*(13), 2078-2091.

Metzger, M.J., Flanagin, A.J., Eyal, K., Lemus, D., & McCann, R. (2003). Bringing the concept of credibility into the 21st century: Integrating perspectives on source, message, and media credibility in the contemporary media environment. *Communication Yearbook, 27*, 293-335.

Moosbrugger, H., & Goldhammer, F. (2007). FAKT-II. Frankfurter Adaptiver Konzentrationsleistungs-Test II. Manual. Grundlegende neu bearbeitete und neu norminierte 2. Auflage des FAKT von Moosbrugger und Heyden (1997). Bern: Huber.

Muthén, B.O. (1998-2004). Mplus Technical Appendices. Los Angeles, CA: Muthén & Muthén.

Muthén, L. K., & Muthén, B. O. (1998-2011). Mplus User's Guide. Sixth Edition. Los Angeles, CA: Muthén & Muthén.

Naumann, J., Richter, T., & Groeben, N. (2001). Validierung des INCOBI anhand eines Vergleichs von Anwendungsexperten und Anwendungsnovizen. *Zeitschrift für Pädagogische Psychologie, 3* (4), 219–232.

Perfetti, C.A. (1985). *Reading ability.* New York: Oxford University Press.

OECD (2011). *PISA 2009 Results: Students on Line: Digital Technologies and Performance* (Volume VI). Available from: http://dx.doi.org/10.1787/9789264112995-en [Accessed 17 October 2012].

R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Rains, S.A., & Karmikel, C.D. (2009). Health information-seeking and perceptions of website credibility: Examining Web-use orientation, message characteristics, and structural features of websites. *Computers in Human Behavior, 25*, 544–553.

Richter, T., Naumann, J., Isberner, M.-J., & Kutzner, Y. (2011). Diagnostik von Lesefähigkeiten bei Grundschulkindern: Eine prozessorientierte Alternative zu produktorientierten Tests [Assessment of reading skills in primary school children: A process-oriented alternative to product-oriented tests]. *Diskurs Kindheits- und Jugendforschung, 6*, 479-486.

Rieh, S.Y. (2002). Judgment of information quality and cognitive authority in the Web. *Journal of the American Society for Information Science and Technology, 53*(2), 145-161.

Rizopoulos, D. (2006). ltm: An R Package for Latent Variable Modeling and Item Response Theory Analyses. *Journal of Statistical Software, 17*(5) 1-25. Available from: http://www.jstatsoft.org/v17/i05/ [Accessed 17 October 2012].

Rouet, J-F., Ros, C., Goumi, A., Macedo-Rouet, M. &, Dinet, J. (2011). The influence of surface and deep cues on primary and secondary school students' assessment of relevance in Web menus. *Learning and Instruction, 21*, 205-219.

Salmerón, L., Kammerer, Y., & García-Carríon, P. (2013). Searching the Web for conflicting topics: Page and user factors. *Computers in Human Behavior, 29*, 2161-2171.

Salmerón, L., Naumann, J., García, V., & Farjado, I. (2017). Scanning and deep processing of information in hypertext: an eye tracking and cued retrospective think-aloud study. *Journal of Computer Assisted Learning, 33*, 222–233.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Test of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research - Online, 8*, 23-74.

Schipolowski, S., Schroeders, U., & Wilhelm, O. (2008). BEFKI - Berlin Test of Fluid and Crystallized Intelligence. Poster presented at the XXIX International Congress of Psychology, Berlin, Germany.

Schwarz, J. & Morris, M. R. (2011). Augmenting Web Pages and Search Results to Support Credibility Assessment. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, May 07-12, 2011, Vancouver, BC, Canada [doi>10.1145/1978942.1979127]

Sutcliffe, A. & Ennis, M. (1998). Towards a cognitive theory of information retrieval. *Interacting with Computers, 10*, 321-351.

Thomm, E., & Bromme, R. (2012). “It Should at Least Seem Scientific!” Textual Features of “Scientificness” and Their Impact on Lay Assessments of Online Information. *Science Education, 96*(2), 187–211.

Tu, Y., Shih, M., & Tsai, C. (2008). Eighth graders’ web searching strategies and outcomes: The role of task types, web experiences and epistemological beliefs. *Computers and Education, 51*, 1142–1153.

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension.* New York: Academic Press.

Van Gog, T., Paas, F., & van Merriënboer, J. J. G. (2005). Uncovering expertise-related differences in troubleshooting performance: combining eye movement and concurrent verbal protocol data. *Applied Cognitive Psychology, 19*, 205-221.

Wallace, R. M., Kupperman, J., Krajcik, J., & Soloway, E. (2000). Science on the web: Students online in a sixth-grade classroom. *The Journal of the Learning Sciences, 9*, 75–104.

Walraven, A., Brand-Gruwel, S., & Boshuizen, H. (2008). Information-problem solving: A review of problems students encounter and instructional solutions. *Computers in Human Behavior, 24*, 623–648.

Walraven, A., Brand-Gruwel, S., & Boshuizen, H. (2009). How students evaluate information and sources when searching the World Wide Web for information. *Computers and Education, 52*, 234–246.

Walraven, A., Brand-Gruwel, S., & Boshuizen, H. (2010). Fostering transfer of websearchers’ evaluation skills: A field test of two transfer theories. *Computers in Human Behavior, 26*, 716–728.

Wathen, C.N. & Burkell, J. (2002). Believe it or not: Factors influencing credibility on the Web. *Journal of the American Society for Information Science and Technology, 53*(2), 134-144.

Wirth, W., Böcking, T., Karnowski, V., & von Pape, T. (2007). Heuristic and systematic use of search engines. *Journal of Computer-Mediated Communication, 12*, 778–800.

Paper 3

Publication Note

Keßel, Y., Goldhammer, F., & Kröhne, U. (under review, *Computers in Human Behavior*). Successful Information Evaluation Online: The Role of Task Characteristics and Information Processing Behavior.

Successful Information Evaluation Online: The Role of Task Characteristics and Information Processing Behavior

Nowadays most people use the web as their predominant information source and often base important personal decisions on the information they find online. Therefore, it has become increasingly important to understand how people make use of search engines in depth and which variables determine a successful information evaluation on the web. Thus, the present study investigated how characteristics of online search results and the way of interacting with these search results are associated with successful information evaluation. Results obtained by the analyses of log-file data showed that a high number of information sources hinder a successful information evaluation, while congruency between credibility features on search engine result pages and corresponding web sites facilitates the evaluation process. When individuals visited a high number of different web sites, the probability of successful information evaluation was increased. However, this effect was less pronounced with a rising number of information sources.

1. 1ntroduction

The use of search engines (e.g., Google, Yahoo!, Bing) has become part of our daily life to deal with the unlimited amount of information available on the World Wide Web (web) and to fulfill our current information needs. The possibility of a directed web search and the great variety of up-to-date information, which is not only easily accessible, but also cost-efficient, has made the web superior to other media (Wittwer, Bromme, & Jucks, 2004). Especially in education, students most often use the web as their only or predominant source of information and rarely revert to traditional information sources (Beljaarts, 2006; Jones, 2002). Despite the obvious advantages of searching information online, there are special risks in the web search process. In contrast to traditional media, online information often does not underlie any editorial review and can easily be altered (Metzger, Flanagin, & Zwarun, 2003; Gerjets, Kammerer & Werner, 2011). Especially the introduction of web 2.0 has led to an enormous increase of information suppliers (Lucassen & Schraagen, 2011). Therefore, the credibility of online information accessed by search engines is highly variable. The ability to critically evaluate online information independently has become indispensable. However, studies have shown that many internet users have difficulties evaluating the credibility of online information adequately (Brand-Gruwel, Wopereis, & Vermetten, 2005; Gerjets & Hellenthal-Schorr, 2008). Even students, being frequent users of the web, display major imperfections evaluating the quality of “search engine result pages” (SERPs) and web sites (Lorenzen, 2002; MaKinster, Beghetto, & Plucker, 2002; Wallace, Kupperman, Krajcik, & Soloway, 2000; Walraven, Brand-Gruwel & Boshuizen, 2009). For instance, Gerjets et al. (2011) have shown that individuals predominantly access the first few web pages presented on a SERP neglecting the fact that even the top search results might not be credible information sources and one-sided or commercially biased due to search engine optimization business (Lewandowski, 2011, 2013). However, research has proven that the users' ability to critically evaluate online information sources can be improved by adequate interventions (e.g. Argelagós & Pifarré, 2012; Braasch, Bråten, Strømsø, Anmarkrud, & Ferguson, 2013; Gerjets & Hellenthal-Schorr, 2008; Mason, Junyent, & Tornatora, 2014; Walraven, Brand-Gruwel, & Boshuizen, 2010). Therefore, it is of great importance to further examine influencing variables and specific processes during web search in order to develop adequate training programs for students and be able to individually foster their evaluation skills.

In this paper we investigated two major factors affecting successful information evaluation: the context, in which the web search task takes place, defined by various characteristics (the so-called "task characteristics"), and the individual person performing the web search task by applying information processing strategies indicated by behavioral process characteristics. Task characteristics refer to properties of the SERP and linked web pages, whereas individual process characteristics refer to individual differences in the search behavior (e.g., visiting a web page). Considering task and individual process characteristics is expected to provide insights into the differences between more and less competent evaluators of online information. The present paper aims at gaining a deeper understanding of the web search process by investigating influencing variables using multiple data sources including task characteristics, log-file data providing process measures as well as outcome measures of the web search task.

2. Theoretical Background

2.1 Evaluation Processes during Web Search

Studying predictors for a successful evaluation of online information requires taking a closer look at evaluation processes during web search activities. The web search process has been described by different models and can be divided into several steps (Brand-Gruwel et al., 2005; Gerjets et al., 2011; Gerjets & Hellenthal-Schorr, 2008; Rouet, Ros, Goumi, Macedo-Rouet, & Dinet, 2011). Five important steps have been identified (adapted from Gerjets et al., 2011, p. 221): As a first step, an information problem is recognized and a search goal is defined. In a second step of a web search process, a search engine (e.g. Google, Yahoo!, Bing) is selected, a search term is entered and sent. Subsequently, a SERP with a list of search result links is returned to the user. In a third step the list of search results is scanned and the links are evaluated regarding their significance for the search goal. Links are considered and then selected for further inspection. At this time, the evaluation of search results is based on sparse information about the corresponding web sites and the information they likely contain (e.g., titles, text fragments, domain names and suffixes). One major observable individual process characteristic is the time spent on each of typically multiple SERPs. The following observable action in the information search process is clicking on a link and entering the corresponding web site. In a fourth step after entering a selected web site, its information is scanned, evaluated with regard to its relevance for the search goal, and, in case of relevance, information is extracted for further processing. Here, another observable process characteristic is the time spent on a web site. The fifth step of the web search process contains comparison and evaluation of different web sites with regard to their credibility. In this last step, the information is integrated towards a solution of the information problem. Particularly the last three steps, namely the evaluation of search results, the evaluation of web sites and their comparison with respect to their credibility, are crucial for evaluation processes during web search (Gerjets et al. 2011). Therefore, the present study focuses on these last three steps. While many studies have concentrated only on investigating evaluation criteria and factors influencing credibility during web search (Crystal & Greenberg, 2006; Fritch & Cromwell, 2001; Rieh, 2002; Savolainen & Kari, 2005; Tombros, Ruthven, & Jose, 2005; Wogalter & Mayhorn, 2008), this study does not only focus on task characteristics, but also explicitly addresses individual process characteristics during the web search that might serve as potential predictors for a successful information evaluation online. In the following, these potentially influencing characteristics are described in detail.

2.2 Criteria Affecting Credibility Judgments during Web Search

2.2.1 Information credibility features.

The evaluation of online information refers to the judgment of many different information features, some of them reflecting the content of information and others reflecting its presentation (Lucassen & Schraagen, 2011). In line with this distinction, Wathen and Burkell (2002) propose a model for the evaluation of online information with two categories: the evaluation of surface credibility based on superficial cues and the evaluation of message credibility based on deep semantic cues (see also Rouet et al., 2011). Similarly, Rains & Karmikel (2009) state that, when judging the credibility of online information, it is essential to investigate message features reflecting the message content by reading the text and/or investigate accompanying information relating to its presentation, namely structural features. Hence, the overall credibility of a web site is indicated by message and correlated structural features. Therefore, they both are important credibility features when searching and evaluating online information (Hong, 2006).

2.2.2 The influence of task characteristics during web search.

Task characteristics determine the context, in which the web search process takes place. They invoke specific search processes differing in their level of complexity. Thus, they are expected to influence an individual's success when evaluating search results online. Dual-processing models such as the Heuristic Systematic Model (HSM) or the Elaboration Likelihood Model (ELM) (Chen & Chaiken, 1999; Petty & Cacioppo, 1986) describe two modes of information processing: systematic processing via a central route and heuristic processing via a peripheral route. Whereas systematic information processing implies a complete evaluation of all given information (e.g., message features such as currency, completeness, objectivity etc.) and therefore requires a larger amount of cognitive resources, heuristic information processing relies on superficial cues (e.g., structural features such as domain name, page rank etc.) that are expected to require less cognitive resources. Whether systematic or heuristic information processing is carried out, is moderated by task characteristics and cognitive ability (Evans, 2008). Thus, task characteristics in general encourage specific ways of information processing during web search, which might rise or reduce the probability of a successful information evaluation.

Several task characteristics influencing the mode of information processing during web search have been studied in the past. Among them are those on a macro level, namely task type and (objective) task complexity, and those on a micro level such as the position of the search result on a SERP (Pan et al., 2007; Salmerón, Kammerer, & García-Carrión, 2013), the number of links returned by the search engine and their characteristics, the complexity and structure of a web site etc. (Wirth, Böcking, Karnowski, & von Pape, 2007). Objective task complexity conceptualized by Campbell (1988) "implies an increase in information load, information diversity, or rate of information change" (p.43) and according to Li and Belkin (2010) is based on the quantity of sub-tasks involved. However, many more conceptualizations of task complexity exist with Liu and Li (2012) giving a very broad definition of task complexity as “the aggregation of any intrinsic task characteristic that influences the performance of a task” (p.559). In this paper we will focus on three task characteristics reflecting task complexity, one of them the number of information sources. According to Wirth & Schweiger (1999) users will more likely use heuristic information processing if the complexity of the decision situation increases. Therefore, it can be assumed that when faced with a small number of information sources (i.e., task complexity is low), a search engine user might invest more time on identifying the most credible information source and rely on more systematic and deep information processing. In consequence this might result in a higher probability of identifying the most credible information source. Of course, actual internet users normally encounter a multiplicity of search results that are returned by the search engine, but generally only a reduced amount of search results is evaluated in terms of relevance credibility. When there are only very few information sources that need to be compared according to their credibility (i.e., the task complexity is low), presumably more cognitive resources are available rather allowing a systematic information processing. Working memory (WM) resources (e.g. Baddeley, 2012) are not needed to process, store and retrieve a great amount of information, but instead can be devoted to carry out a deep information processing – again rising the probability of identifying the most credible information source. According to Baddeley (2012) WM is in charge for temporarily storing information and at the same time controlling information processing. It has been discussed to play an important role in hypermedia learning and has also been explored in relation to navigational processing in hypermedia environments (e.g., Kornmann et al., 2016; Naumann, Richter, Christmann, & Groeben, 2008; Niederhauser, Reynolds, Salmen, & Skolmoski, 2000).

Task complexity is also affected by the degree of attractiveness of the different information sources that need to be compared. We define the attractiveness of an information source by the total amount of credibility features (message and structural features) that indicate high credibility of the source. Consequently, the more credibility features indicate high credibility, the higher the attractiveness of an information source. When less credible and therefore less valuable information sources are relative attractive, they only differ in very few credibility features from the most credible source. The higher the overall attractiveness of the less credible information sources, the more difficult it should be to identify the most credible source. Task complexity is high and a raised level of cognitive effort and WM including executive processes such as focusing attention on relevant information cues while inhibiting irrelevant information are necessary in order to solve the information problem successfully. The number of dimensions of the information requiring attention (i.e., information load) (Schroder, Driver, & Streufert, 1967) and ambiguity of the task (Braarud & Kirwan, 2011) are high due to the many similar credibility features resulting in high task complexity and reducing task success. In addition, due to limited cognitive capacity internet users might rely on heuristic information processing (Wirth & Schweiger, 1999), although in this specific situation, heuristic information processing might lead to false conclusions because the minimal differences between credible and non-credible sources might not be detected properly.

Another task characteristic reflecting task complexity and being expected to influence a successful evaluation of online information is the congruency between credibility features (message and structural features) present on the SERP and those present on the corresponding web site. Both might be unidirectional, but they might also differ. For instance, a link on a SERP might seem promising because its credibility features assure high quality, whereas the corresponding web site contains credibility features identifying the web site as a less credible information source. It should be easier to successfully identify the most credible information source, when credibility features of a link on the SERP and of the web site are congruent and both indicate high or low credibility because less cognitive resources are required and there is no conflict between components of the task. However, dealing with conflicting and controversial or multi-perspective information requires a large amount of cognitive resources such as a heightened level of concentration and WM capacity (Kornmann et al., 2016). Thus, only relying on the information credibility features present on the SERP and not visiting the corresponding web sites, would lead to false conclusions regarding the credibility of a search result. Research has shown that although web users seem to be aware of the importance of information credibility when searching for controversial topics online, they still heavily rely on heuristic information processing strategies as for example the “top link” heuristic (Pan et al., 2007; Salmerón et al., 2013), which might be leading to an insufficient information evaluation (cf. Rouet et al., 2011). Therefore, a successful evaluation of online information should be easier when credibility features of a link on a SERP and the corresponding web site are congruent.

2.2.3 The influence of individual process characteristics during web search.

Complementary to task characteristics providing the context in which the web search takes place and therefore requiring certain ways of information processing, the web search process is also influenced by characteristics of the individual carrying out of the web search. As people differ in their ability to successfully evaluate online information, they can be expected to carry out search and evaluation processes differently. Thus, it is assumed that competent web searchers are more likely to show a specific web search behavior. To be able to explore search and evaluation processes further empirically, it is important to first identify traces of the web search process by investigating log-file-data. In a second step indicators can be derived from these traces of task completion, which provide important information about the individual search processes. During web search activities there are several observable actions reflecting the process of web search, which can be used to construct process indicators. One of the observable actions is clicking on a link on a SERP and accessing the corresponding web site. Gerjets et al. (2011) consider this step the central action in the information search process. It provides information about which web sites are accessed and which information is available later in the process when different information sources are compared and evaluated with respect to their credibility. Thus, the process of evaluating online information is in part reflected by an indicator that is defined as the number of visited web sites. Granka, Feusner and Lorigo (2008) show that individuals normally access only three to four search results per SERP. Search engine users also tend to predominantly select and access the first few search results on a SERP (Guan & Cutrell, 2007; Joachims, Granka, Pan, Hembrooke, & Gay, 2005). However, this search strategy has obvious flaws, as ranking algorithms of most search engines are mainly based on page relevance and popularity thus reflecting information quality not directly (Cho & Roy, 2004). Consequently, web sites listed among the first search results on a SERP may contain subjective, unscientific, commercially biased and even wrong information due to search engine optimization and the so-called richer-get-richer effect (i.e., highly ranked pages are predominantly selected by users, receive more links and are in turn ranked higher by the search engine) (Kammerer & Gerjets, 2014, Lewandowski, 2013). Therefore, it can be supposed that accessing a higher number of web sites, especially not only just the first two or three search results but also search results ranked lower, is highly beneficial for a successful information evaluation, because it raises the probability of accessing credible web sites. However, since internet users usually only have limited capacity of time and effort they will invest during a web search process, applying suitable heuristics can be a matter of importance.

Time-related information can be used to extract further important indicators providing information about the web search process itself. For instance, the total time that is spent on SERPs as well as the total time that is spent on web sites offers valuable clues about how intensively information and credibility features on SERPs and Web sites are studied. Initial source evaluations on SERPs are essential in the web search process because from the large number of web pages usually classified as relevant by a search engine, only a few can be inspected in greater detail (cf. Braasch et al., 2009; Rieh 2002). Hence, investing time to evaluate the information presented on SERPs instead of just rapidly clicking through the result list, should result in a higher probability of accessing web sites containing credible information and therefore foster a successful evaluation process. Similarly, spending more time on the selected web sites should increase the probability of identifying a credible information source. The content of the web sites and the connected credibility features can be studied more intensively, similarly to the deep and differentiated information processing via the central route in the ELM as shown by Petty and Cacioppo (1986). This should facilitate storing valuable credibility cues in memory and retrieving them when in a last step of the web search process the information from different sources is compared with respect to its credibility.

Furthermore, studies have shown that spending more time on a task is beneficial for task success in particular, when individuals with low ability complete difficult tasks requiring controlled processing (Goldhammer et al., 2014). Hence, it can be assumed that spending more time on SERPs and web sites fosters a successful information evaluation in general, and it is particularly important for individuals with a low ability to evaluate online information.

2.3 Research Hypotheses

The overall goal of this study was to investigate how the web search and evaluation process is related to the probability of successfully evaluating online information as offered by search engines. To address this goal, both task characteristics and individual process characteristics were considered with regard to their prediction of task success. First, task characteristics were assumed to affect task success. Task characteristics evoke specific processing strategies such as heuristic or systematic processing with a certain error-rate in consequence affecting task success. More specifically, the extent to which task characteristics differing in their level of complexity determine task difficulty was investigated. Second, individual process characteristics that served as predictors for a successful evaluation of online information were identified.

The first three hypotheses address task characteristics representing the context of the web search process. They refer to the task attributes attractiveness of non-target search results, number of information sources and congruency of credibility features, which are assumed to have an impact on task difficulty. (1) High attractiveness of non-target search results has a negative effect on the successful task completion, that is, the task is more difficult. (2) When internet users are confronted with a high number of information sources this is expected to have a negative effect on the probability of task success, that is, the task is more difficult. (3) Congruency between credibility features of the SERP and linked web pages has a positive effect on the probability of task success, that is, the task is expected to be less difficult.

The following three hypotheses refer to the individual process characteristics assumed to have an impact on a successful evaluation of online information. (4) We assume that the number of visited web sites offered on a SERP has a positive effect on the successful completion of the task. In case more web sites are visited, the higher the probability of task success. Visiting more web sites results in a greater diversity of information that can be evaluated and compared and therefore reduces the probability of missing highly credible information pieces, that might not be displayed on the SERP but found on a web site. Consequently, visiting more web sites rises the probability of identifying a credible information source. (5) The time spent on a SERP is assumed to have an effect on successful task completion. The more time is spent on a SERP, the higher the probability of task success. Investing more time to investigate the links on a SERP increases the quality of information processing because it allows a more thorough evaluation of information and therefore increases the probability of identifying a credible information source. (6) The time spent on the different web sites linked on a SERP is assumed to have an effect on successful item completion. The more time is spent on the different Web sites, the higher the probability of a thorough, in-depth and complete processing of all given information, which is more likely to result in task success.

3. Material and Methods

3.1 Participants

Participants were 379 students in grade 9 from all three German school tracks (49% female, 51% male) at several schools in Germany, with an overall mean age of 16.69 (*SD* = .67) ranged from 15 to 19 years.

3.2 Materials

3.2.1 Test for evaluating online information.

The cognitive ability to judge the credibility of online information using structural features and message features efficiently was assessed by the subscale "Evaluating Structural and Message Features" (ESMF) from the "Test for Evaluating Online Information" (TEO). The TEO test consists of 24 interactive tasks simulating web environments.

The subscale ESMF consists of eight tasks with online information presented in the form of Google-like SERPs and corresponding web sites (cf. Table 1). The tasks contain either three or five search result entities on one SERP and a corresponding web site for each of the search results. Structural and message features indicating credibility or non-credibility are available for evaluation. The corresponding linked web sites are static HTML sites with all hyperlinks, buttons or menus being inactive. The online information included in the items consists of factual texts, which can be considered either to be true or false. Participants are asked to identify the most credible link/ web site by selecting one of the radio buttons placed beside each of the search result entities. For each task, the individual response and the response behavior were recorded. There is only one correct response. The correct response is defined by the link/ web site with the highest number of information features indicating credibility. Selecting the correct response is scored with 1, selecting a false response is scored with 0. Indicators of the response process such as times spent on pages (measured in seconds) and frequencies of web site visits were extracted from log-file data. Three indicators were composed: The number of visited web sites was generated by accumulating the number of different visited web sites within one task. The time spent on a SERP was defined as the accumulated time (log-transformed and measured in seconds) an individual spent on the SERP within one task including revisits of the SERP from the task call up to the task response. The time spent on web sites was defined as the accumulated time (log-transformed and measured in seconds) spent on all web sites within one task including revisits of web sites from the call up of the first web site to the task response. A typical task for the subscale ESMF implemented with the Item Builder software is displayed in Figure 1 (Rölke, 2012).

Figure 1. Example Screenshot of a SERP from the ESMF scale with three search results.

For all tasks, the positions of the correct and false responses on the SERP were random. The tasks of the ESMF subscale were designed taking task characteristics into account affecting the task difficulty. The following attributes are balanced across the tasks: the attractiveness of non-target search results (high vs. low), the congruency (yes vs. no) of structural features (superficial cues present on a SERP) and message features (deep cues present in the content of the Web sites) and the number of information sources to be compared (three vs. five). If the attractiveness of the non-target search result entities on a SERP is high compared to the target search result, they only differ in very few structural and/or message features indicating differences in credibility, and therefore, high attractive non-target search result entities are more difficult to distinguish from the target search result. If the attractiveness of the non-target search results is low compared to the target search result, the search results differ in many structural and/or message features facilitating the distinction. Attractiveness of non-target search results and congruency of structural and message features (yes vs. no) were verified based on expert ratings.

The congruency of structural and message features was varied using one group of (congruent) tasks with structural and message features both either indicating credibility or non-credibility and another group of (incongruent) tasks with structural features signaling credibility and message features signaling non-credibility. Concerning the number of information sources (three vs. five) a task was assumed to be more difficult, if many information sources need to be compared. The search topics health, crafts, sports and education were also balanced across tasks. They were not expected to have an impact on task difficulty.

Table 1. Item Properties of the Subscale ESMF from the TEO.

	Dimensions of Task Design			Context Factor	
Task	Attractiveness of Non-Targets	Congruency of Structural and Message Features	Number of Online Sources to be evaluated	Task Topic	Position of Target link on the SERP
1	High	Yes	5	Sports	3
2	High	Yes	3	Education	2
3	High	No	5	Health	4
4	High	No	3	Crafts	2
5	Low	Yes	5	Education	3
6	Low	Yes	3	Health	3
7	Low	No	5	Crafts	2
8	Low	No	3	Health	3

Reliability analysis revealed a reliability of .50 (Cronbach's α value) for the subscale ESMF of the TEO. Items proportion of correct responses ranged between .28 and .64, with six of the eight tasks showing a proportion of correct responses under .50. That means tasks were rather difficult. Item discriminations were good and ranged between .39 and .56. Items fit a unidimensional Rasch-model.

3.3 Procedure

Data was collected in the context of a pilot study of the *National Educational Panel Study* (NEPS) funded by the Federal Ministry of Education and Research in Germany. Tests were administered in groups of up to 24 students in computer labs at schools. The subscale ESMF of the TEO was administered at the beginning of the test sessions. After the TEO tests, students also completed a questionnaire and some other tests, however, these tests were not of importance for the present study. The overall completion time was about 120 minutes with a 15-minute break after 60 minutes.

3.4 Data Analysis

Analyses are based on dichotomous task responses (selected entity on the SERP) and the three indicators number of visited web sites, time spent on a SERP and time spent on web sites extracted from the log-file data.

To investigate how task characteristics and individual process characteristics predict a successful information evaluation, regression analyses within the Generalized Linear Mixed Modelling (GLMM) framework were conducted by means of the lme4-package (Bates, Maechler, Martin, & Bolker, 2011) in the R-environment (R Development Core Team, 2011). Regression analyses included item covariates (task characteristics) and person-by-item-covariates (individual process variables) and were carried out estimating fixed effects for predictors and intercepts varying randomly across items (item easiness) and persons (skill level), respectively. Furthermore, possible interactions between task characteristics and individual process characteristics were explored with the help of nested model comparisons. To test if the two models differ significantly, a model difference test (ANOVA) was conducted. For the statistical tests of significance an alpha level of .05 was assumed.

4. Results

4.1 Results on Hypothesis 1, 2 and 3 Investigating the Effect of Task Characteristics on Task Success

To investigate if the task characteristics attractiveness of non-targets, number of information sources and congruency of credibility features show a significant effect on task success, that is, a successful evaluation of online information, a generalized linear mixed model was tested with all three task characteristics as predictors and item-level scores as the predicted criterion. Results of the multiple regression analysis can be found in Table 2.

Table 2. Multiple Regression Model with Task Characteristics.

Predictor	b	SE(b)
Attractiveness of Non-Targets	-.02	.24
Number of Information Sources	-.54*	.24
Congruency of Credibility Features	.59*	.24

Note: ***p < .001. **p < .01. *p < .05; SE(b) = Standard error of regression weight b.

As shown in Table 2, Hypothesis 1 on the effect of attractiveness of non-targets was not supported by the results. The task characteristic attractiveness of non-targets did not reveal a significant (positive) effect on task success. In contrast, Hypotheses 2 and 3 were clearly supported by the data. The number of information sources had a significant negative effect on task success, whereas for the congruency of credibility features a significant positive effect on task success was revealed (cf. Table 2).

4.2 Results on Hypothesis 4, 5 and 6 Investigating the Effect of Individual Process Characteristics on Task Success

To test, if the individual process characteristics number of visited web sites, time spent on a SERP and time spent on web sites show a significant effect on the successful information evaluation online, a multiple regression analysis with the mentioned variables as predictors and the item-level scores as predicted criterion was conducted (Table 3).

Table 3. Multiple Regression Model with Individual Process Characteristics.

Predictor	b	SE(b)
Number of Visited web sites	.44***	.03
Time Spent on SERP	.10	.10
Time Spent on web sites	.03	.03

Note: ***p < .001. **p < .01. *p < .05. SE(b) = Standard error of regression weight b.

As shown in Table 3, the number of visited web sites had a highly significant positive effect on task success. Thus, successful evaluation of online information was easier when a high number of web sites were visited. In conclusion, Hypothesis 4 was supported by the data. In Hypothesis 5 it was assumed that the time spent on a SERP has a significant positive effect on task success. However, the process variable time spent on a SERP did not show a significant effect on task success (cf. Table 3). Thus, Hypothesis 5 was not supported by the data. In Hypothesis 6 it was claimed, that the time spent on web sites has a significant positive effect on task success. Yet, as shown in Table 3, there was no significant effect revealed for the time spent on web sites. Thus, neither Hypothesis 6 was supported by the data.

4.3 Exploration of Interaction Effects between Task Characteristics and Individual Process Characteristics

Given that task and individual process characteristics jointly affect task success we furthermore explored, whether and how task and individual process characteristics interact with each other. Thus, interaction analyses were conducted to clarify whether the effect of individual process characteristics depends on task characteristics and vice versa.

To investigate possible interaction effects, first a complete multiple regression analysis with all task and individual process characteristics was conducted. Results for the multiple regression model can be found in Table 1 of the appendix. As before, significant predictors at task level proved to be the number of information sources as well as the congruency of credibility features, and at individual level the number of visited web sites.

In a second step, interactions between task characteristics and individual process characteristics were investigated by conducting nine separate multiple regression analyses for each possible interaction effect. Subsequently, for each regression model incorporating an interaction term, a model comparison test (ANOVA) has been conducted to investigate whether adding the interaction term to the regression model substantially improves the model fit. An overview of the results for all model comparison tests can be found in Table 2 of the appendix.

Regression analysis revealed only one significant interaction effect that substantially improves the model fit (χ^2=5.96, *df*=1, *p*<.05). This was a significant negative interaction for the task processes characteristic number of information sources and the individual processes characteristic number of visited web sites. The negative interaction means that if a person completes a task presenting many information sources, the number of visited web sites is less crucial for task success. The final multiple regression model including this significant interaction effect can be found in Table 4.

Table 4. Multiple Regression Model with Task, Individual Process Characteristics and Interaction.

Predictor	b	SE(b)
Attractiveness of Non-Targets	-.03	.28
Number of Information Sources	-.81**	.30
Congruency of Credibility Features	.91**	.28
Number of Visited web sites	.57***	.06
Time Spent on SERP	.11	.10
Time Spent on web sites	.03	.03
Interaction: Number of Information Sources × Number of Visited websites	-.17*	.07

Note: ***p < .001. **p < .01. *p < .05. SE(b) = Standard error of regression weight b.

5. Discussion and Conclusions

We investigated the impact of task characteristics such as the attractiveness of non-targets, the number of information sources that need to be compared and the congruency of structural credibility features (superficial credibility cues on a SERP) and message features (deep credibility cues reflecting the content of a web site) on the successful evaluation of information. Furthermore, we wanted to identify individual process variables representing individual solution behavior in the web search process and use them as predictors for a successful evaluation of online information. Based on our assumption that individual process variables extracted from log-files can be understood as empirical correlates of cognitive information processing, we investigated the effect of the number of visited web sites, the time spent on SERPs and the time spent on linked web sites. In addition, we examined possible interaction effects between task characteristics and individual process characteristics.

5.1 The Role of Task Characteristics

The results revealed a significant negative effect on task success for the number of information sources. Accordingly and as expected, having to compare a relatively high number of information sources, that is, links on SERPs and web sites, reduces the probability of identifying the most credible information source. This effect might be due to web users applying information processing strategies depending on external factors such as the complexity of information as well as other situational characteristics of the decision situation itself like the number of decision alternatives (Wirth et al., 2007). In decision situations of relatively high uncertainty and high subjective task complexity, individuals typically aim at maximizing the outcome and minimizing the cognitive effort (Gigerenzer & Goldstein, 1996). Thus, in decision situations that require choosing between a large number of alternatives (e.g., a high number of different information sources), individuals might need to apply heuristic processing rules like the "top link" heuristic (Pan et al., 2007) instead of carrying out a thorough and systematic evaluation, because the latter exceeds the individual cognitive resources (e.g., working memory capacity). In case of applying only the "top link" heuristic though, this behavior might lead to false conclusions concerning information credibility as top search results can be commercially biased or one-sided (Lewandowski, 2011).

Furthermore, the task characteristic congruency of credibility features showed a significant positive effect on task success. Tasks with congruent credibility features in search results on SERPs and on the corresponding web sites seem to be easier than tasks with incongruent credibility features. Again, this effect could be due to a limited working memory capacity. When credibility features between links on SERPs and web sites are not congruent, more and especially conflicting information has to be processed and stored requiring high memory capacity and making it more difficult to inhibit irrelevant information. In addition, the conflicting information needs to be integrated towards a solution for the information problem requiring high working memory capacity.

Interestingly, the task characteristic attractiveness of non-targets did not show a significant effect on the successful evaluation of information. This might be due to the task design. Although the search results that have to be compared are very similar, the small number of search results that need to be compared theoretically allows a quite thorough and deep evaluation, which can be supposed to result in correct choices of the most credible information source. Although additional statistical analyses did not show an interaction between the number of search results and the attractiveness of non-targets, our assumption was still supported by analyses which revealed that users spent significantly more time on tasks with high attractiveness of non-targets than on tasks with low attractiveness of non-targets presumably applying deep and systematic evaluation strategies. This might be different with a larger number of search results as circumstances of high uncertainty and limited time and capacity usually result is the use of heuristic rules (Wirth & Schweiger, 1999), which might then rather lead to false conclusions. For future research it would therefore be interesting to test the effect of the attractiveness of non-targets on task success with tasks comprehending larger sets of search results.

5.2 The Role of Individual Process Characteristics

Results for the individual process characteristics showed that the number of visited web sites has a significant positive effect on task success. Thus, the more different web sites are visited, the higher is the probability of identifying the most credible information source. But the data also revealed that more is not always better. In contrast to our expectations results neither showed a significant effect for the time spent on SERPs nor for the time spent on web sites. Studies have shown that individuals are not always completely aware of the evaluation criteria they apply or simply do not use known evaluation criteria when conducting a web search (Fogg et al, 2003). Quick and unconscious evaluation processes also affect the selection when trying to identify the most credible information source. Therefore, it might be possible that some individuals spontaneously and unconsciously apply the right evaluation criteria while others need more time to consciously reflect on the evaluation process. This assumption is supported by additional analyses that do not show a significant positive effect on task success for the individual total time spent on the ESMF tasks.

5.3 Interactions of Task Process Characteristics and Individual Process Characteristics

When investigating possible interactions between task characteristics and individual process characteristics, we found a significant negative interaction for the number of information sources and the number of visited web sites. The more information sources need to be evaluated and compared, the less predictive is the number of visited web sites for task success. Hence, when there are only a few information sources that need to be evaluated and compared individuals benefit from using a highly elaborated information processing strategy (systematic processing) considering all alternatives in depth, that is visiting and exploring every web site. With higher numbers of information sources though this systematic strategy becomes less helpful because it may come at a cost of mental effort as it requires the use of a large amount of cognitive resources and at some point exceeds the available cognitive capacity (Rieh, Kim & Markey, 2012). Therefore, in these situations of high uncertainty due to the large number of decision alternatives, a rather heuristic evaluation maximizing the outcome but minimizing time and cognitive effort may be more beneficial (Gigerenzer & Goldstein, 1996; Metzger, 2007; Taraborelli, 2008; Wirth et al., 2007).

5.4 Limitations and Future Research

Findings of the present study are based on insights obtained by the application of the TEO consisting of tasks simulating the web and web search situations. Therefore, the present study has certain limitations that should be acknowledged when interpreting the results. Tasks of the TEO represent only an approximation of a standardized web and web search process. The ESMF scale of the TEO consists of tasks with three to five search results. Of course, actual search engines will always return a vast amount of search results to the user. However, although a real web search provides thousands of results, users usually focus on the first few search results/ search results pages (Cutrell & Guan, 2007; Kammerer & Gerjets, 2014, Pan et al., 2007) and only compare a small number of search results (web sites) in the end. Therefore, we believe that the TEO tasks represent a limited but highly standardized approach of the search process and adequately reflect this stage of the web search process despite their small number of search results. In addition, TEO tasks only represent a small section of the web search process. The first two steps of the web search process according to Gerjets et al. (2011) are not incorporated in the tasks. Thus, individuals do not have to recognize an information problem and define an adequate search term, which nowadays that search engines use auto completion lists and therefore already give feedback to the user during the entry of the search term, can already be seen as a part of the evaluation process. Due to the web that is constantly changing and developing new features and functions, the results are restricted to the technology used in the tasks and surely have to be replicated with tasks incorporating up-to-date technologies.

Furthermore, a limitation of this study is that the results are restricted to 15-year-old students in grade 9 of the German school system. Results might be different for other age groups. Moreover, a rather small number of tasks was used to investigate the cognitive processes underlying the credibility evaluation of online information. Effects might be stronger when using more tasks. Therefore, for future research the TEO subscale should be extended by the development of additional tasks.

The present study does not claim to be exhaustive in its selection of task characteristics and individual process characteristics. The log-file data we used does not describe and represent all possible processes carried out during a web search. Especially, evaluation processes that go beyond overt actions remain largely undiscovered. To unravel these processes future research should incorporate further methodological approaches. For instance, eye-tracking would allow a reconstruction of every search result on a SERP that was looked at irrespective from its selection. Van Gog, Paas and Van Merriënboer (2005) advise to combine eye-tracking and thinking-aloud protocols to gain deeper insights into cognitive processes; therefore, the latter should be applied as well in future research.

For future research it would also be interesting to investigate navigation paths during web search and identifying ideal navigation paths. This way, the distance between ideal and empirical navigation paths could be described and used as an individual process characteristic giving us an even deeper insight into the process of web search. As has been studied an ideal navigation path fostering comprehension and in consequence a successful information evaluation might contain more linear sequencing, less clicking backwards and a more coherent navigation (Naumann et al., 2008; Richter, Naumann, Brunner, & Christmann, 2005; Salmerón & Garcia, 2011). Furthermore, navigation paths can also play a role when investigating the effect of task characteristics on task success, as they might mediate the relationship between task characteristics and the probability of task success. In case of incongruent credibility features in links on the SERP and on the corresponding web sites navigation behavior such as subsequently entering each web site in contrast to not navigating at all can be assumed to mediate task success. Therefore, it should be rewarding to investigate mediation models as well. In addition, now that we have explored the effect of task and individual process characteristics on a successful information evaluation online, it also seems of interest to extend the analyses by investigating possible interactions with user characteristics such as domain knowledge, web experience, motivation, working memory capacity or epistemic beliefs that have already been shown to affect the process of evaluating of online information (Kammerer & Gerjets, 2012; Kammerer, Aman, & Gerjets, 2015; Kornmann et al., 2016; Metzger, 2007).

We believe that the present findings contribute to a deeper and differentiated understanding of the process of credibility evaluations on the web and hope that the limitations and possible improvements outlined will advance future research providing further insights into this complex process fostering a conscious and competent handling of the diversity of information found online.

References

Argelagós, E., & Pifarré, M. (2012). Improving information problem solving skills in secondary education through embedded instruction. *Computers in Human Behavior, 28*, 515–526.

Baddeley, A. D. (2012). Working memory: theories, models, and controversies. *Annual Review of Psychology, 63*, 1-29. http://dx.doi.org/10.1146/annurevpsych-120710-100422.

Bates, D., Maechler, Martin & Bolker, B. (2011). *lme4: Linear mixed-effects models using S4 classes.* R package version 1.1-13. http://CRAN.R-project.org/package=lme4

Beljaarts, M. (2006). *Internet, een populair medium voor het zoeken van informatie bij schoolopdrachten [Internet, a popular medium for searching information for school assignments]*. 's-Hertogenbosch: Malmberg.

Braarud, P.Ø., Kirwan, B., 2011. Task complexity: what challenges the crew and how do they cope. In: Skjerve, A.B., Bye, A. (Eds.), *Simulator-Based Human Factors Studies Across 25 Years: the History of the Halden Man-Machine Laboratory*. Springer, London, pp. 233-251.

Braasch, J. L. G., Lawless, K. A., Goldman, S. R., Manning, F. H., Gomez, K. W., & MacLeod, S. M. (2009). Evaluating search results: an empirical analysis of middle school students' use of source attributes to select useful sources. *Journal of Educational Computing Research, 41*, 63–82.

Braasch, J. L. G., Bråten, I., Strømsø, H. I., Anmarkrud, Ø., & Ferguson, L. E. (2013). Promoting secondary school students' evaluation of source features of multiple documents. *Contemporary Educational Psychology, 38*, 180–195.

Brand-Gruwel, S., Wopereis, I., & Vermetten, Y. (2005). Information problem solving by experts and novices: analysis of a complex cognitive skill. *Computers in Human Behavior,* 21, 487-508.

Campbell, D. J. (1988). Task complexity: A review and analysis. *Academy of Management Review, 13*(1), 40–52.

Chen, S., & Chaiken, S. (1999). The heuristic-systematic model in its broader context. In S. Chaiken & Y. Trope (Eds.). *Dual-Process Theories in Social Psychology* (pp. 73-96). New York: Guilford Press.

Cho, J., & Roy, S. (2004). Impact of search engines on page popularity. In: S. Feldman and M. Uretsky, Eds. *Proceedings of the 13th International Conference on World Wide Web* (WWW '04). New York, NY: ACM Press, 20–29.

Crystal, A., & Greenberg, J. (2006). Relevance criteria identified by health information users during web searches. *Journal of the American Society for Information Science and Technology, 57*, 1368-1382.

Cutrell, E., & Guan, Z. (2007). What are you looking for? An eye-tracking study of information usage in Web search. In M. B. Rosson & D. Gilmore (Eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '07; pp. 407–416). New York, NY: ACM Press.

Evans, J. S. (2008). Dual-processing accounts of reasoning, judgment and social cognition. *Annual Review of Psychology, 59*, 255-278.

Fogg, B. J., Soohoo, C., Danielson, D. R., Marable, L., Stanford, J., & Tauber, E. R. (2003). How do users evaluate the credibility of Web sites? A study with over 2,500 participants. In *Proceedings of the 2003 conference on designing for user experiences (DUX '03)* (pp. 1–15). New York, NY: ACM Press.

Fritch, J.W., & Cromwell, R.L. (2001). Evaluating Internet resources: Identity, affiliation, and cognitive authority in a networked world. *Journal of the American Society for Information Science and Technology, 52*(6), 499-507.

Gerjets, P., & Hellenthal-Schorr, T. (2008). Competent information search in the World Wide Web: development and evaluation of a web training for pupils. *Computers in Human Behavior, 24*, 693-715.

Gerjets, P., Kammerer, Y., & Werner, B. (2011). Measuring spontaneous and instructed evaluation processes during Web search: Integrating concurrent thinking-aloud protocols and eye-tracking data. *Learning and Instruction, 21*, 220-231.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*, 650–669.

Goldhammer, F., Naumann, J., Stelter, A., Tóth, K., Rölke, H., Klieme, E. (2014). The time on task effect in reading and problem solving is moderated by task difficulty and skill. *Journal of Educational Psychology, 106* (3), 608-626.

Granka, L., Feusner, M., & Lorigo, L. (2008). Eye monitoring in online search. In R. I. Hammoud (Ed.), *Passive eye monitoring: Algorithms, applications and experiments* (pp. 283-304). Berlin: Springer.

Guan, Z., & Cutrell, E. (2007). An eye tracking study of the effect of target rank on Web search. In M. B. Rosson & D. J. Gilmore (Eds.), *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 417-420). New York: ACM Press.

Hong, T. (2006). The influence of structural and message features on website credibility. *Journal of the American Society for Information Science and Technology, 57*(1), 114-127.

Joachims, T., Granka, L., Pan, B., Hembrooke, H., & Gay, G. (2005). Accurately interpreting click through data as implicit feedback. In R. Baeza-Yates, N. Ziviani, G. Marchionini, A. Moffat, & J. Tait (Eds.), *Proceedings of the 28th annual internationalACMSIGIR conference on research and development in information retrieval* (pp. 154-161). New York: ACM Press.

Jones, S. (2002). *The Internet Goes to College. How Students are Living in the Future with Today's Technology.* Washington, DC: Pew Internet and American Life Project.

Kammerer, Y., & Gerjets, P. (2012). Effects of search interface and Internet-specific epistemic beliefs on source evaluations during Web search for medical information: An eye-tracking study. *Behaviour & Information Technology, 31*, 83–97.

Kammerer, Y., & Gerjets, P. (2014). The Role of Search Result Position and Source Trustworthiness in the Selection of Web Search Results When Using a List or a Grid Interface. *International Journal of Human-Computer-Interaction, 30*, 177-191.

Kammerer, Y., Amann, D.G., & Gerjets, P. (2015). When adults without university education search the Internet for health information: The roles of Internet-specific epistemic beliefs and a source evaluation intervention. *Computers in Human Behavior, 48*, 297-309.

Kornmann, J., Kammerer, Y., Anjewierden, A., Zettler, I., Trautwein, U., & Gerjets, P. (2016). *Computers in Human Behavior, 55*, 145-158.

Lewandowski, D. (2011). The influence of commercial intent of search results on their perceived relevance. In S. Sadagopan & K. Ramamritham (eds.). *Proceedings of the 2011 iConference* (iConference '11; pp. 452-458). New York, NY: ACM Press.

Lewandowski, D. (2013). Credibility in Web search engines. In S. Apostel & M. Folk (Eds.), *Online credibility and digital ethos: Evaluating computer-mediated communication* (pp. 131-146). Hershey, PA: IGI Global.

Li, Y., & Belkin, N. J. (2010). An exploration of the relationships between work task and interactive information search behavior. *Journal of the American Society for Information Science and Technology, 61*(9), 1771–1789.

Liu,P., & Li, Z. (2012). Task complexity: A review and conceptualization framework. *International Journal of Industrial Ergonomics, 42*, 553-568.

Lorenzen, M. (2002). The land of confusion? – High school students and their use of the World Wide Web for research. *Research strategies, 18*, 151-163. Retrieved September 7, 2006 from http://www.libraryinstruction.com/confusion.html

Lucassen, T., & Schraagen, J.M. (2011). Factual accuracy and trust in information: The role of expertise. *Journal of the American Society for Information Science and Technology, 62*(7), 1232-1242.

MaKinster, J. G., Beghetto, R. A., & Plucker, J. A. (2002). Why can't I find Newton's third law? Case studies of students'use of the web as a science resource. *Journal of Science Education and Technology, 11*, 155-172.

Mason, L., Junyent, A. A., & Tornatora, M. C. (2014). Epistemic evaluation and comprehension of web-source information on controversial science-related topics: Effects of a short-term instructional intervention. *Computers & Education, 76*, 143–157.

Metzger, M. (2007). Making Sense of Credibility on the Web: Models for Evaluating Online Information and Recommendations for Future Research. *Journal of the American Society for Information Science and Technology, 58*(13), 2078-2091.

Metzger, M.J., Flanagin, A.J., & Zwarun, L. (2003). College student Web use, perceptions of information credibility, and verification behavior. *Computers & Education, 41*, 271-290.

Naumann, J., Richter, T., Christmann, U., & Groeben, N. (2008). Working memory capacity and reading skill moderate the effectiveness of strategy training in learning from hypertext. *Learning and Individual Differences, 18*, 197-213.

Niederhauser, D. S., Reynolds, R. E., Salmen, D. J., & Skolmoski, P. (2000). The influence of cognitive load on learning from hypertext. *Journal of Educational Computing Research, 23*, 237-255.

Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G., & Granka, L. (2007). In Google we trust: Users' decisions on rank, position and relevance. *Journal of Computer-Mediated Communication, 12*, 801-823.

Petty, R. E. & Cacioppo, J. T. (1986). The Elaboration Likelihood Model of Persuasion. In: *Advances in experimental social psychology* (Ed. L. Berkowitz), 19, pp. 123 – 205. New York: Academic Press.

R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Rains, S.A. & Karmikel, C.D. (2009). Health information-seeking and perceptions of website credibility: Examining Web-use orientation, message characteristics, and structural features of websites. *Computers in Human Behavior, 25*, 544–553.

Richter, T., Naumann, J., Brunner, M., & Christmann, U. (2005). Strategische Verarbeitung beim Lernen mit Text und Hypertext [Strategic Processing in Learning with Text and Hypertext]. *German Journal of Educational Psychology, 19*, 5-22.

Rieh, S. Y. (2002). Judgment of information quality and cognitive authority in the web. *Journal of the American Society for Information Science and Technology, 53*, 145-161.

Rieh, S. Y., Kim, Y. M., & Markey, K. (2012). Amount of invested mental effort (AIME) in online searching. *Information Processing and Management, 48*(6), 1136–1150.

Rölke, H. (2012). The item builder: A graphical authoring system for complex item development. In T. Bastiaens & G. Marks (Eds.), *Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education* (pp. 344–353). Chesapeake, VA: AACE. Retrieved October 24, 2013, from http://www.editlib.org/p/41614.

Rouet, J-F., Ros, C., Goumi, A., Macedo-Rouet, M. &, Dinet, J. (2011). The influence of surface and deep cues on primary and secondary school students' assessment of relevance in Web menus. *Learning and Instruction, 21*, 205-219.

Salmerón, L., & García, V. (2011). Reading skills and children's navigation strategies in hypertext. *Computers in Human Behavior, 27*, 1143-1151.

Salmerón, L., Kammerer, Y., & García-Carrión, P. (2013). Searching the Web for conflicting topics: Page and user factors. *Computers in Human Behavior, 29*, 2161-2171.

Savolainen, R., & Kari, J. (2005). User-defined relevance criteria in web searching. *Journal of Documentation, 62*, 685-707.

Schroder, H.M., Driver, M.J., & Streufert, S. (1967). *Human Information Processing: Individuals and Groups Functioning in Complex Social Situations.* Holt, Rinehart and Winston, New York, NJ, USA.

Taraborelli, D. (2008). How the web is changing the way we trust. In A. Briggle, K. Waelbers, & P. A. E. Brey (Eds.), *Current issues in computing and philosophy* (pp. 194–204). Amsterdam, the Netherlands: IOS Press.

Tombros, A., Ruthven, I., & Jose, J. M. (2005). How users assess web pages for information-seeking. *Journal of the American Society for Information Science and Technology, 56*, 327-344.

Van Gog, T., Paas, F., & Van Merriënboer, J. J. G. (2005). Uncovering expertise-related differences in troubleshooting performance: combining eye movement and concurrent verbal protocol data. *Applied Cognitive Psychology, 19*, 205-221.

Wallace, R. M., Kupperman, J., Krajcik, J., & Soloway, E. (2000). Science on the web: students online in a sixth-grade classroom. *The Journal of the Learning Sciences, 9*(1), 75-104.

Walraven, A., Brand-Gruwel, S., & Boshuizen, H. P. A. (2009). How students evaluate sources and information when searching the World Wide Web for information. *Computers & Education, 52*(1), 234-246.

Walraven, A., Brand-Gruwel, S., & Boshuizen, H. P. A. (2010). Fostering transfer of websearchers' evaluation skills: A field test of two transfer theories. *Computers in Human Behavior, 26*, 716–728.

Wathen, C.N., & Burkell, J. (2002). Believe it or not: Factors influencing credibility on the Web. *Journal of the American Society for Information Science and Technology, 53*(2), 134-144.

Wirth, W., & Schweiger, W. (1999). Selektion neu betrachtet: Auswahlentscheidungen im Internet [Selection revisited: Selection decisions on the Internet]. In W. Wirth & W. Schweiger (Eds.), *Selektion im Internet [Selection on the Internet]* (pp. 43–74). Opladen, Germany: Westdeutscher Verlag.

Wirth, W., Böcking, T., Karnowski, V. & von Pape, T. (2007). Heuristic and systematic use of search engines. *Journal of Computer-Mediated Communication, 12*, 778-800.

Wittwer, J., Bromme, R., & Jucks, R. (2004). Kann man dem Internet trauen, wenn es um die Gesundheit geht? Die Glaubwürdigkeitsbeurteilung medizinischer Fachinformationen im Internet durch Laien. *Zeitschrift für Medienpsychologie, 2*, 48-56.

Wogalter, M.S., & Mayhorn, C.B. (2008). *International Journal of Technology and Human Interaction, 4*(1), 75-93.

APPENDIX

1. Erklärung zur Promotionsordnung
2. Eidesstattliche Versicherung
3. Erklärung zu bisherigen Promotionsverfahren
4. Stellungnahme zu den Kriterien einer publikationsbasierten Dissertation
5. Erklärung über die Eigenleistung
6. Bestätigung der Einreichung
7. Lebenslauf

Erklärung zur Promotionsordnung

Ich erkläre hiermit, dass mir die Promotionsordnung der Mathematisch-Naturwissenschaftlichen Fachbereiche der Goethe Universität Frankfurt am Main bekannt ist.

Frankfurt am Main, den 04. Oktober 2017

Yvonne Keßel

Eidesstattliche Versicherung

Ich erkläre hiermit, dass ich die vorgelegte Dissertation mit dem Titel *Development of Interactive Performance Measures for two Components of ICT Literacy: Successfully Accessing and Evaluating Information* selbständig angefertigt und mich nicht anderer Hilfsmittel als der in ihr angegebenen bedient habe. Insbesondere versichere ich, dass alle Entlehnungen aus anderen Schriften mit Angabe der betreffenden Schrift gekennzeichnet sind. Ich habe die Grundsätze der guten wissenschaftlichen Praxis beachtet und habe nicht die Hilfe einer kommerziellen Promotionsvermittlung in Anspruch genommen.

Frankfurt am Main, den 04. Oktober 2017

Yvonne Keßel

Erklärung zu bisherigen Promotionsverfahren

Ich erkläre hiermit, dass ich mich bisher keiner Doktorprüfung im Mathematisch-Naturwissenschaftlichen Bereich unterzogen habe.

Frankfurt am Main, den 04. Oktober 2017

Yvonne Keßel

Stellungnahme zu den Kriterien einer publikationsbasierten Dissertation

Kriterien für kumulative Dissertationen im Fachbereich Psychologie und Sportwissenschaften, Goethe Universität Frankfurt (nach Beschluss im Fachbereichsrat gültig ab 11.06.2015):

(1) Die kumulative Dissertation soll in der Regel 3 Schriften umfassen, die aus den letzten 5 Jahren stammen sollen.

Erklärung: Die Dissertation umfasst eine Schrift aus dem Jahr 2013, die vorab am 20. Oktober 2012 erstmals online publiziert wurde, sowie zwei derzeit in Fachzeitschriften eingereichte Manuskripte.

(2) Die Schriften sollen im Wesentlichen einem zusammenhängenden Forschungsprogramm entstammen. Die jeweils verfolgten Forschungsfragen sollen sich sinnvoll zueinander in Beziehung setzen lassen.

Erklärung: Die Schriften sind alle mit der Entwicklung und Validierung interaktiver und computerbasierter Testverfahren befasst, welche die Kompetenzen zur Beschaffung und Bewertung von Informationen im Internet erfassen.

(3) Der Kandidat oder die Kandidatin soll bei 2 Publikationen Erstautor/Erstautorin sein, bei einer weiteren Publikation kann er/sie Koautor/Koautorin sein. Eine geteilte Erstautorenschaft wird für jeden der Erstautoren anteilig gewichtet (bei 2 Erstautoren eine 1/2 Erstautorenschaft, bei 3 eine 1/3 Erstautorenschaft usw.).

Erklärung: Yvonne Keßel ist alleinige Erstautorin bei zwei Manuskripten. Es gibt keine geteilten Erstautorenschaften. Bei einem Manuskript ist sie Koautorin.

(4) Die drei Schriften sollen zur Veröffentlichung zumindest eingereicht sein. Der aktuelle Status ist detailliert darzulegen (Publikationsorgan und Status wie eingereicht, in revision, conditional accept usw.).

(5) Mindestens 2 der 3 Schriften müssen in guten oder sehr guten, in der Regel englischsprachigen, Zeitschriften mit Peer-Review eingereicht sein.

(6) Eine der 3 Schriften kann als Publikation in einem einschlägigen Lehrbuch, Enzyklopädieband oder einem anderen für das jeweilige Fach bedeutsamen Publikationsorgan, jeweils mit Peer-Review, eingereicht oder veröffentlicht sein.

Erklärung zu (4)-(6): Eine bereits publizierte Schrift ist bei dem *European Journal of Psychological Assessment* (Impact Factor 2016: 2.328) erschienen. Eine weitere Schrift wurde am 03.10.2017 bei der Fachzeitschrift Interacting with Computers (Impact Factor 2016:

1.410) eingereicht. Die dritte Schrift wurde am 03.10.2017 bei der Fachzeitschrift Computers in Human Behavior (Impact Factor 2016: 3.435) eingereicht. Alle Schriften sind englischsprachig. Die genannten Fachzeitschriften arbeiten mit einem Peer-Review-Verfahren.

(7) Die als Dissertation vorgelegte Abhandlung soll über die zusammengestellten Publikationen hinaus einen zusätzlichen Text enthalten, in welchem eine kritische Einordnung der eigenen Publikationen aus einer übergeordneten Perspektive heraus vorgenommen wird. Dieser Text sollte einen Umfang von ca. 30 Seiten haben. Es sollen die Fragestellungen theoretisch entwickelt werden, die empirischen Arbeiten und ihre Ergebnisse so dargestellt werden, dass sie auch ohne Lesen der Einzelarbeiten nachvollziehbar sind und es soll eine Gesamtdiskussion enthalten, die die Fragestellungen beantwortet und den Erkenntnisgewinn der Arbeit herausstellt.

(8) Die Dissertation muss eine Erklärung enthalten, in der die Eigenleistung des Kandidaten/der Kandidatin dargestellt wird. Insbesondere bei Schriften mit Koautoren, aber auch bei in Einzelautorenschaft entstandenen Schriften, die oft auch im Rahmen von Abteilungsprojekten, Drittmittelprojekten, Projektverbünden usw. entstanden sind, soll dargelegt werden, welchen Anteil die Kandidaten an Entwicklung der Fragestellung, Design, Durchführung, Auswertung der empirischen Studie(n) und an dem Abfassen der einzelnen Beiträge hatten. Diese Erklärung ist von Betreuer und/oder Koautoren zu bestätigen.

Erklärung: Die entsprechenden Texte sind enthalten.

(9) In besonders begründeten Fällen kann von diesen Richtlinien abgewichen werden.
(10) Bei den vorgeschlagenen Kriterien handelt es sich um Empfehlungen. Es wird explizit darauf hingewiesen, dass natürlich nach wie vor die jeweilige Promotionsordnung, die Beschlüsse des Promotionsausschusses und die von den Gutachtern erstellten Gutachten entscheidend für das Verfahren sind.

Anmerkung: Satz (8) gilt auch für Dissertationen, die als Monographie vorgelegt werden.

Frankfurt am Main, den 04. Oktober 2017

Yvonne Keßel

Erklärung über die Eigenleistung

Die vorliegende Dissertation mit dem Titel *Development of Interactive Performance Measures for two Components of ICT Literacy: Successfully Accessing and Evaluating Information* beinhaltet drei Manuskripte, die zum Zeitpunkt der Eröffnung des Promotionsverfahrens in internationalen Fachzeitschriften veröffentlicht oder zur Veröffentlichung eingereicht worden sind. Die in Beitrag I dieser Dissertation beschriebene Studie basiert auf Daten, die im Rahmen des deutschen Feldversuchs des *Programme for International Student Assessment (PISA)* 2009 erhoben wurden. Die Daten der Studien in Beitrag II und III dieser Dissertation wurden im Kontext der *National Educational Panel Study (NEPS)* in den *Mode Effect Entwicklungsstudien* A37 im Jahr 2010 sowie A36 im Jahr 2011 erhoben. Das Projekt wurde gefördert vom *Bundesministerium für Bildung und Forschung (BMBF)*. Die Verfasserin der Dissertation, Yvonne Keßel, arbeitete als wissenschaftliche Mitarbeiterin von Mai 2009 bis August 2015 am Deutschen Institut für Internationale Pädagogische Forschung (DIPF) im Zentrum für technologiebasiertes Assessment (TBA-Zentrum). Sie war an der computerbasierten Implementation neuer und an der Überarbeitung bereits bestehender Testverfahren sowie an der Datenaufbereitung und Dokumentation maßgeblich beteiligt bzw. dafür verantwortlich.

Beitrag I: Assessing Individual Differences in Basic Computer Skills
Goldhammer, F., Naumann, J. & **Keßel, Y**. (2013). *European Journal of Psychological Assessment, 29* (4), 263-275.

Das Manuskript wurde bei der Zeitschrift *European Journal of Psychological Assessment* am 20.10.2012 online veröffentlicht. Dieser Beitrag stellt einen neu entwickelten Test zur Erfassung basaler Computerfähigkeiten (Basic Computer Skills Test, BCS) dar. Die Entwicklung der Testaufgaben, Datenauswertung und Erstellung des Manuskriptentwurfes wurden hauptverantwortlich vom Erstautor durchgeführt. Yvonne Keßel war insbesondere mit der inhaltlichen Analyse und Überarbeitung der Testaufgaben befasst. Des Weiteren gab sie ausführliche inhaltliche sowie sprachliche Rückmeldungen zum Manuskript und arbeitete an der Revision mit.

Beitrag II: Evaluating the Credibility of Online Information: The Influence of Basic Computer Skills, Word Recognition and Reasoning
Keßel, Y., Goldhammer, F., & Kröhne, U. (in Begutachtung, *Interacting with Computers*).

Das Manuskript wurde am 03.10.2017 zur Begutachtung bei der Fachzeitschrift *Interacting with Computers* eingereicht. Die Fragestellungen dieses Beitrags entwickelte die Erstautorin hauptverantwortlich in Diskussion mit ihrem Betreuer, Prof. Dr. Frank Goldhammer. Der Test zur Erfassung basaler Computerfähigkeiten (BCS Test) stellt eine Weiterentwicklung des in Beitrag I verwendeten Tests dar. Die BCS-Testaufgaben entwickelte die Erstautorin in Abstimmung mit Prof. Dr. Frank Goldhammer weiter und setzte diese mit dem CBA Itembuilder eigenständig um. Zudem ergänzte sie den Test durch die Entwicklung weiterer neuer Testaufgaben grundlegend. Die Testaufgaben des Tests zur Bewertung von Online Informationen (Test for Evaluating Online Information, TEO) entwickelte und erstellte die Erstautorin selbstständig mit dem CBA Itembuilder. Die im Manuskript berichteten Datenanalysen wurden von der Erstautorin mit der Statistik-Software Mplus eigenständig durchgeführt und mit ihrem Betreuer diskutiert. Das Manuskript wurde vollständig von der Erstautorin verfasst. Die Ko-Autoren haben erst und ausschließlich vor der Einreichung Überarbeitungsvorschläge eingebracht.

Beitrag III: Successful Information Evaluation Online: The Role of Task Process Characteristics and Information Processing Behavior

Keßel, Y., Goldhammer, F., & Kröhne, U. (in Begutachtung, *Computers in Human Behavior*).

Das Manuskript wurde am 03.10.2017 bei der Fachzeitschrift *Computers in Human Behavior* zur Begutachtung eingereicht. Der Beitrag entstand in Anknüpfung an die Analysen des zweiten Beitrags. Die Fragestellung wurde hauptverantwortlich von der Erstautorin in Diskussion mit Prof. Dr. Frank Goldhammer entwickelt. Die Aufgaben des TEO Tests wurden von der Erstautorin für diesen Beitrag eigenständig überarbeitet und angepasst. Die im Manuskript berichteten Datenanalysen wurden von der Erstautorin mit der Statistik-Software R eigenständig durchgeführt und anschließend mit ihrem Betreuer diskutiert. Die Erstautorin war federführend für das Abfassen des Manuskriptes verantwortlich. Die Ko-Autoren haben erst und ausschließlich vor der Einreichung Überarbeitungsvorschläge eingebracht.

Frankfurt am Main, den 04. Oktober 2017

Dipl.-Psych. Yvonne Keßel	Prof. Dr. Frank Goldhammer
Verfasserin der Dissertation	Betreuer der Dissertation

Bestätigung der Einreichung (1/2)

Beitrag II: **Keßel, Y.**, Goldhammer, F., & Kröhne, U. (under review, *Interacting with Computers*). Evaluating the Credibility of Online Information: The Influence of Basic Computer Skills, Word Recognition and Reasoning.

ScholarOne Manuscripts https://mc.manuscriptcentral.com/iwc

Interacting with Computers

Home

Author

Review

Submission Confirmation

Print

Thank you for your submission

Submitted to	Interacting with Computers
Manuscript ID	IWC-17-0139
Title	Evaluating the Credibility of Online Information: The influence of basic computer skills, word recognition and reasoning
Authors	Keßel, Yvonne Goldhammer, Frank Kröhne, Ulf
Date Submitted	03-Oct-2017

Author Dashboard

Bestätigung der Einreichung (2/2)

Beitrag III: **Keßel, Y.**, Goldhammer, F., & Kröhne, U. (under review, *Computers in Human Behavior*). Successful Information Evaluation Online: The Role of Task Characteristics and Information Processing Behavior

Von: eesserver@eesmail.elsevier.com Gesendet: Di 03.10.2017 17:49
An: y.kessel@arcor.de
Cc:
Betreff: Submission Confirmation

Computers in Human Behavior
Title: Successful Information Evaluation Online: The Role of Task Characteristics and Information Processing Behavior
Authors: Yvonne Nicole Keßel; Frank Goldhammer, Prof. Dr.; Ulf Kröhne, Dr.
Article Type: Full Length Article

Dear YvonneKessel,

Your submission entitled "Successful Information Evaluation Online: The Role of Task Characteristics and Information Processing Behavior" has been received by Computers in Human Behavior.

I will send the paper out for two peer reviews and will be getting back to you as soon as possible with a decision.

You may check on the progress of your paper by logging on to the Elsevier Editorial System as an author. The URL is https://eeslive.elsevier.com/chb/.

Your username is: y.kessel@arcor.de

If you need to retrieve password details, please go to: http://ees.elsevier.com/CHB/automail_query.asp.

Your manuscript will be given a reference number once an Editor has been assigned.

Thank you for submitting your work to this journal. Please do not hesitate to contact me if you have any queries.

Kind regards,

Computers in Human Behavior

Lebenslauf

Yvonne Nicole Keßel geb. Pfaff
geboren am 24.01.1983
in Frankfurt am Main

Adresse: Wingertstraße 24
60316 Frankfurt am Main

Tel.: 069-20737287
Email: y.kessel@arcor.de

Berufliche Qualifikation

seit 10/2010	Selbstständige Tätigkeit in eigener Praxis für Psychologische Beratung und Lösungsorientierte Kurzzeittherapie, Frankfurt
05/2009 - 08/2015	Wissenschaftliche Mitarbeiterin zur Promotion am Deutschen Institut für Internationale Pädagogische Forschung (DIPF) im Zentrum für technologiebasiertes Assessment (TBA, Leiter: Prof. Dr. Frank Goldhammer)

Berufsausbildung

10/2009 -09/2010	Ausbildung in Integrierter Lösungsorientierter Psychotherapie und Coaching an der ILP-Fachschule, Mainz (Abschluss: 1,1)

Schul- und Hochschulausbildung

seit 07/2009	Promotionsstudium an der Goethe-Universität Frankfurt am Main im Fachbereich Psychologie unter der Betreuung von Prof. Dr. Frank Goldhammer
10/2002 – 02/20009	Studium der Psychologie an der Goethe-Universität, Frankfurt am Main, Abschluss: Diplom (1,2) Thema der Diplomarbeit: Ein LST-Modell zum situationsspezifischen Ärgerausdruck

1993 – 2002 Gymnasium, Philipp-Reis-Schule, Friedrichsdorf
Abschluss: Abitur (1,8)

Zeitschriftenbeiträge mit Peer-Review Verfahren

Goldhammer, F., Naumann, J., & **Keßel, Y**. (2013). *European Journal of Psychological Assessment, 29* (4), 263-275.

Goldhammer, F., Kröhne, U., **Keßel, Y**., Senkbeil, M., & Ihme, J.M. (2014). Diagnostik von ICT-Literacy: Multiple-Choice- vs. simulationsbasierte Aufgaben [Assessment of ICT literacy: Multiple-choice vs. simulation-based tasks]. *Diagnostica, 60*, 10-21.

Einreichungen

Keßel, Y., Goldhammer, F., & Kröhne, U. (under review, *Interacting with Computers*). Evaluating the Credibility of Online Information: The Influence of Basic Computer Skills, Word Recognition and Reasoning.

Keßel, Y., Goldhammer, F., & Kröhne, U. (under review, *Computers in Human Behavior*). Successful Information Evaluation Online: The Role of Task Process Characteristics and Information Processing Behavior

Vorträge

Pfaff, Y., & Goldhammer, F. (2011, 01. September). *Measuring individual differences in ICT literacy: Accessing and Evaluating Information.* Vortrag auf der European Association for Research on Learning and Instruction (EARLI) in Exeter, GB.

Pfaff, Y., & Goldhammer, F. (2011, 14. September). *Messung individueller Unterschiede im Bereich ICT Literacy: Bewertung von Online Informationen.* Vortrag auf der 13. Fachgruppentagung Pädagogische Psychologie der DGPs (PAEPS) in Erfurt, DE.

Keßel, Y., Goldhammer, F., Magmet, V., & Kröhne, U. (2012, 25. September). *Informationsbewertung im Internet - Die Rolle von Prozessvariablen als Prädiktoren für den Aufgabenerfolg.* Vortrag auf dem 48. Kongress der Deutschen Gesellschaft für Psychologie in Bielefeld, DE.