

Naumann, Johannes; Goldhammer, Frank; Rölke, Heiko; Stelter, Annette  
**Erfolgreiches Problemlösen in technologiereichen Umgebungen.  
Wechselwirkungen zwischen Interaktionsschritten und  
Aufgabenanforderungen**

*formal und inhaltlich überarbeitete Version der Originalveröffentlichung in:  
formally and content revised edition of the original source in:*

*Zeitschrift für pädagogische Psychologie 28 (2014) 4, S. 193-203*



Bitte verwenden Sie in der Quellenangabe folgende URN oder DOI /  
Please use the following URN or DOI for reference:  
urn:nbn:de:0111-pedocs-148308  
10.25656/01:14830

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-148308>  
<https://doi.org/10.25656/01:14830>

#### Nutzungsbedingungen

Dieses Dokument steht unter folgender Creative Commons-Lizenz:  
<http://creativecommons.org/licenses/by-nc/4.0/deed.de> - Sie dürfen das  
Werk bzw. den Inhalt vervielfältigen, verbreiten und öffentlich zugänglich  
machen sowie Abwandlungen und Bearbeitungen des Werkes bzw. Inhaltes  
anfertigen, solange Sie den Namen des Autors/Rechteinhabers in der von ihm  
festgelegten Weise nennen und das Werk bzw. den Inhalt nicht für  
kommerzielle Zwecke verwenden.  
Mit der Verwendung dieses Dokuments erkennen Sie die  
Nutzungsbedingungen an.

#### Terms of use

This document is published under following Creative Commons-License:  
<http://creativecommons.org/licenses/by-nc/4.0/deed.en> - You may copy,  
distribute and render this document accessible, make adaptations of this work  
or its contents accessible to the public as long as you attribute the work in the  
manner specified by the author or licensor. You are not allowed to make  
commercial use of the work, provided that the work or its contents are not  
used for commercial purposes.  
By using this particular document, you accept the above-stated conditions of  
use.



#### Kontakt / Contact:

peDOCS  
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation  
Informationszentrum (IZ) Bildung  
E-Mail: [pedocs@dipf.de](mailto:pedocs@dipf.de)  
Internet: [www.pedocs.de](http://www.pedocs.de)

Mitglied der

  
Leibniz-Gemeinschaft

## Akzeptierte Manuskriptfassung (nach peer review) des folgenden Artikels:

Naumann, J., Goldhammer, F., Rölke, H. & Stelter, A. (2014). Erfolgreiches Problemlösen in technologiebasierten Umgebungen: Wechselwirkungen zwischen Interaktionsschritten und Aufgabenanforderungen. *Zeitschrift für Pädagogische Psychologie*, 28 (4), 193-203.

doi: 10.1024/1010-0652/a000134

© Hogrefe Verlag, Bern 2014

Diese Artikelfassung entspricht nicht vollständig dem in der Zeitschrift veröffentlichten Artikel. Dies ist nicht die Originalversion des Artikels und kann daher nicht zur Zitierung herangezogen werden.

Die akzeptierte Manuskriptfassung unterliegt der Creative Commons License CC-BY-NC.

**J. Neumann et al.: Problemlösen in technologiebasierten Umgebungen**

**Erfolgreiches Problemlösen in technologiebasierten Umgebungen:  
Wechselwirkungen zwischen Interaktionsschritten und  
Aufgabenanforderungen**

Johannes Naumann<sup>1</sup>, Frank Goldhammer<sup>1,2</sup>, Heiko Rölke<sup>1</sup> und Annette Stelter<sup>1</sup>

<sup>1</sup>Deutsches Institut für Internationale Pädagogische Forschung (DIPF)

<sup>2</sup>Zentrum für internationale Bildungsvergleichsstudien (ZIB)

**Zusammenfassung:** Wir zeigen anhand von Feldtestdaten des “Programme for the International Assessment of Adult Competencies” (PIAAC) in Deutschland ( $N = 661$ ) und Kanada ( $N = 411$ ), dass sich der Erfolg beim Lösen technologiebasierter Probleme aus der Anzahl der Interaktionen mit dem System vorhersagen lässt. Dabei war der Zusammenhang umgekehrt U-förmig. Optimale Lösungsraten resultierten bei einer überdurchschnittlichen Anzahl von Interaktionen. Weiterhin wurde der Zusammenhang durch Merkmale der Aufgabe moderiert. Es resultierte vor allem bei Aufgaben, die eine hohe Zahl von Navigationsschritten erfordern, ein stark positiver Zusammenhang. Entgegen den Erwartungen resultierte bei wohl definierten Aufgaben ein höherer Zusammenhang zwischen

der Anzahl von Interaktionen und der Lösungsrate als bei schwach definierten Aufgaben. Die Ergebnisse aus der deutschen und kanadischen Stichprobe waren parallel.

**Schlüsselwörter:** PIAAC, Computer, Problemlösen, Prozessdaten, Aufgabeneigenschaften

### **Successful Problem Solving in Technology Rich Environments: Interactions Between Number of Actions and Task Demands**

**Abstract:** Using Field Trial data from the Programme for the International Assessment of Adult Competencies (PIAAC) from the German ( $N = 661$ ) and Canadian ( $N = 411$ ) Field Trials, we show that success in technology-based problem solving is predicted by the number of actions taken during problem solving. The association was reversely U-shaped. The probability of succeeding on a problem-solving task was highest at an above-average number of actions. The association between actions taken and problem solving was further moderated by task characteristics. Especially in tasks involving long navigation sequences a strong positive association between number of actions and task success was obtained. Against expectations, the association of actions and probability of task success was higher in well-defined than in ill-defined tasks. Results for the German and the Canadian samples ran parallel.

**Keywords:** PIAAC, Computers, Problem solving, Process data, Item characteristics

## **1 Einleitung**

Dass beim Umgang mit Technologie mit unschöner Regelmäßigkeit „Probleme“ auftreten, ist ein Gemeinplatz. Bis zum Ende des vergangenen Jahrhunderts war die Programmierung eines Videorekorders sprichwörtlich. Zum Zeitpunkt, zu dem wir diesen Aufsatz verfassen, mag eher ein *Smartphone* als Beispiel dienen, das aus unbekanntem Gründen keinen Zugang zu einem *W-LAN* findet. Die Bedeutung von „Problem“ ist hier nah an der in der wissenschaftlichen Psychologie üblichen Definition: Ein Problem liegt dann vor, wenn sich eine Aufgabe nicht ausschließlich unter Einsatz von Routineoperationen lösen lässt. Mit der digitalen Revolution ist für Individuen und Gesellschaften das Lösen von Problemen, die sich aus der Verfügbarkeit und Nutzung von Informations- und Kommunikationstechnologien ergeben, zu einer zentralen Herausforderung geworden (OECD, 2012). Derartige Probleme bezeichnen wir ab hier als „technologiebasierte Probleme“, das entsprechende Problemlösen als „technologiebasiertes Problemlösen“ (vgl. Rammstedt, 2013). Einen Ansatz zur Messung der Kompetenz Erwachsener beim technologiebasierten Problemlösen hat in jüngerer Zeit die OECD mit dem *Programme for the International Assessment of Adult Competencies* (PIAAC) unternommen (OECD, 2013a). Aufgaben zum technologiebasierten Problemlösen aus diesem Assessment bezeichnen wir im Folgenden als TbP- (*Technologiebasiertes Problemlösen*) Aufgaben. Wir wollen in diesem Aufsatz zeigen, dass sich der Erfolg beim technologiebasierten Problemlösen aus einem einfachen Indikator des Problemlöseverhaltens vorhersagen lässt: der Anzahl von Interaktionen mit dem technologischen System (im Folgenden: Anzahl von Interaktionen). Wir leiten diese Hypothese zum einen aus der Literatur zu informationsbezogenen technologiebasierten Problemen ab, zum anderen aus Überlegungen und Befunden zu Explorationsverhalten beim Umgang mit Computertechnologie. Hieraus ergeben sich auch Vorhersagen zur Form des Zusammenhangs und zu Wechselwirkungen zwischen dem Prozessmerkmal „Anzahl von Interaktionen“ und Eigenschaften des

technologiebasierten Problems, insbesondere dem Definitionsgrad der Aufgabe und der Länge des benötigten Navigationspfades.

### **1.1 Technologiebasiertes Problemlösen in PIAAC**

In jüngster Zeit wurde technologiebasiertes Problemlösen als eine Kompetenzdomäne im Rahmen der PIAAC-Studie (OECD 2012; 2013a, b) erfasst. Bei PIAAC handelt es sich um eine in 24 Staaten und Regionen durchgeführte Vergleichsstudie an 16-65-Jährigen (der Bevölkerung im erwerbsfähigen Alter). Erfasst wird die Ausprägung von drei Kompetenzen: Lesefähigkeit, Alltagsmathematische Kompetenz, sowie technologiebasiertes Problemlösen. Technologiebezogene Probleme werden im Kontext von PIAAC unter drei Perspektiven betrachtet: Der Aufgabenstellung, den involvierten Technologien (Hard- und Software sowie Repräsentationsformate) und den beteiligten kognitiven Dimensionen (dem Planen und Überwachen der Problemlösung sowie dem Beschaffen, Bewerten und Nutzen relevanter Informationen; vgl. im Einzelnen OECD, 2012).

Technologiebasierte Probleme im Sinne von PIAAC sind vorwissensreiche – im Sinne der Einleitung zu diesem Themenheft (Greiff, Kretzschmar & Leutner, 2014): domänenspezifische – Probleme. Dies gilt aus zwei Gründen: Erstens müssen die Probandinnen<sup>1</sup> bei PIAAC Kenntnisse über die grundlegende Funktion von Software besitzen, und es müssen in einem Mindestumfang Prozeduren zur Verwendung von Technologie repräsentiert sein. Zweitens handelt es sich bei digitalen Technologien um Technologien zum Verarbeiten symbolisch kodierter Information. Hieraus folgt, dass zur Problemlösung Wissen über die Bedeutung der jeweiligen Symbole vorliegen muss. Die in PIAAC verwendeten Probleme sind statisch in dem Sinne, dass das Problem durch das

---

<sup>1</sup> Um einen guten Lesefluss zu gewährleisten, verwenden wir in diesem Artikel ausschließlich die weibliche Form. Gemeint sind aber immer beide Geschlechter.

jeweilige Item fixiert ist (vgl. z. B Funke, 2006, Öllinger & Knoblich, 2006). Allerdings sind die in PIAAC betrachteten technologiebasierten Probleme komplex (vgl. Greiff et al., 2014) in dem Sinne, dass für die Aufgabenlösung benötigte Information nicht in allen Fällen von vorneherein vollständig gegeben ist, sondern durch Interaktion mit dem Stimulus hergestellt bzw. in einer komplexen Informationsumgebung aufgefunden werden muss. Die in PIAAC thematischen technologiebasierten Probleme variieren darin, ob es sich um – in OECD-Terminologie – wohl definierte (*well-defined*) oder schwach definierte (*ill-defined*) Probleme handelt. Während bei ersteren entsprechend der PIAAC-Nomenklatur ein klar erkennbarer Lösungsweg vorhanden ist, ist es bei letzteren Aufgabe der Testperson, einen Lösungsweg zu finden (OECD, 2012, S. 46). Auch unterscheiden sich die in PIAAC relevanten Probleme dahingehend, wie lang der benötigte Lösungsweg ist, und wie viel Information verarbeitet werden muss.

## **1.2 Die Anzahl von Interaktionen als Prozessindikator**

Wir wollen in dieser Arbeit zeigen, dass sich der Erfolg beim Lösen technologiebasierter Probleme aus einem sehr einfachen Prozessindikator, der Anzahl von Interaktionen, vorhersagen lässt. Insgesamt vermuten wir einen positiven Zusammenhang mit einer erfolgreichen Aufgabenbearbeitung. Wir vermuten, dass es in alltäglichen technologiebasierten Problemen eher ein zu passiver als ein übertrieben aktiver Zugang ist, der einer erfolgreichen Lösung im Wege steht. Wie z.B. aus der Literatur zu unterschiedlichen Trainingsformaten beim Erlernen von Computeranwendungen bekannt ist, zeitigen vor allem solche Trainings gute Ergebnisse, bei denen die Probandinnen ermutigt werden, Lösungswege auszuprobieren und intensiv mit der technologischen Umgebung zu interagieren – auch auf die Gefahr hin, Fehler zu begehen (Bell & Kozlowski, 2008; Dormann & Frese, 1994; Keith & Frese, 2008). Klar ist, dass ein solches Lösungsverhalten

zu einer hohen Anzahl von Interaktionen führt. Trainings, bei denen die Probandinnen instruiert werden, Fehler zu vermeiden, sind dabei deutlich weniger effektiv. Dies gilt insbesondere dann, wenn die Trainingsinhalte auf neue Aufgaben transferiert werden müssen, die Aufgaben also zumindest im Sinne der PIAAC-Terminologie Problemcharakter haben (Keith, Richter, & Naumann, 2010). Die Annahme, dass ein generell zurückhaltender Zugang und entsprechend sparsame Interaktion beim Umgang mit technologiereichen Umgebungen eher hinderlich als zielführend ist, wird auch durch Befunde zum Zusammenhang zwischen Computerängstlichkeit sowie niedriger computerbezogener Selbstwirksamkeit und Performanz gestützt. Diese Beziehung ist generell negativ (für einen Überblick siehe Powell, 2013). Eine Erklärung für diesen Befund könnte darin liegen, dass sich computerängstliche Personen bzw. Personen mit niedriger computerbezogener Selbstwirksamkeit beim Umgang mit dem Computer tendenziell Fehler vermeidend und passiv verhalten. Während im Grundsatz eine positive Beziehung zwischen der Anzahl von Interaktionen und der Lösungswahrscheinlichkeit beim Bearbeiten von TbP-Aufgaben angenommen werden kann, ist diese Beziehung wahrscheinlich nicht linear. Ab einer bestimmten (tendenziell hohen) Anzahl von Interaktionen wird vermutlich ein Punkt erreicht, ab dem zusätzliche Interaktionsschritte kein zielgerichtetes Lösungsverhalten mehr indizieren, sondern eher Desorientierung anzeigen. Hieraus ergibt sich insgesamt die Hypothese eines negativ U-förmigen Zusammenhangs zwischen der Anzahl von Interaktionen und der Lösungswahrscheinlichkeit.

Konsistent mit dieser Vermutung ist die einschlägige Literatur zur Informationssuche in Hypertexten. Die Informationssuche in Hypertexten konstituiert ein (domänenspezifisches) Probleme in dem Sinne, dass entsprechende Aufgaben häufig nicht allein auf der Basis von bereits repräsentierten Schemata gelöst werden können (z.B. Brand-Gruwel, Wopereis & Walraven, 2009, vgl. auch die Problemdefinition bei Greiff et al.,

2014). Bei der Informationssuche in Hypertexten findet man häufig, dass sehr hohe Anzahlen von Interaktionsschritten auf häufige Rücksprünge zurückführbar sind, die für sich genommen negativ mit Performanz assoziiert sind und einen Indikator für Desorientierung darstellen (z. B. Herder & Juvina, 2004; Richter, Naumann, Brunner & Christmann, 2005). Auch findet man auf dem Gebiet der Informationssuche in Hypertexten die hier für TbP-Aufgaben angenommene nichtlineare, negativ beschleunigte und damit tendenziell umgekehrt U-förmige Beziehung zwischen der Anzahl von Interaktionen und Performanz (z.B. OECD, 2011).

Ein solcher nicht-linearer Trend ist dabei allerdings wahrscheinlich nicht das einzige, was den Zusammenhang zwischen der Anzahl von Interaktionen und dem Erfolg beim technologiebasierten Problemlösen charakterisiert. Insbesondere bei Problemen, die längere Navigationssequenzen erfordern, bei denen also häufig zwischen unterschiedlichen Teilen des Stimulus gewechselt werden muss, ist ein stark positiver Zusammenhang zu vermuten. Wenn dagegen nur wenige Navigationsschritte zu bewältigen sind, wird der Zusammenhang vergleichsweise schwach ausfallen.

Außerdem ist minimale Interaktion wahrscheinlich vor allem bei – in PIAAC-Terminologie – schwach definierten Aufgaben schädlich, die Explorationsverhalten erfordern, weil für die Lösung relevante Informationen in der Aufgabenstellung fehlen. Beim Lösen einer TbP-Aufgabe wird sich aufgabenadäquates Explorationsverhalten in einer hohen Anzahl von Interaktionen niederschlagen. Wir gehen also – zusammenfassend – davon aus, dass technologiebasierte Probleme umso schwieriger werden, je mehr Navigationsschritte zu bewältigen sind, und umso schwächer die Aufgabe definiert ist. Beide schwierigkeitsgenerierenden Merkmale können durch geeignete Interaktion mit der technologischen Umgebung kompensiert werden, wobei auch bei schwach definierten technologiebasierten Problemen, die eine hohe Zahl von Navigationsschritten erfordern, ein

Punkt erreicht wird, ab dem mit zusätzlichen Interaktionen die Lösungswahrscheinlichkeit wieder fällt, weil diese zusätzlichen Interaktionen eher Desorientierung als zielgerichtetes Lösungsverhalten anzeigen.

### **1.3 Hypothesen**

Folgende formale Hypothesen zu den in PIAAC verwendeten TbP-Aufgaben lassen sich aus den bis hierher angestellten Überlegungen ableiten: Für das Aufgabenmerkmal Zahl benötigter Navigationsschritte ist ein negativer Zusammenhang mit der Lösungswahrscheinlichkeit zu erwarten (Hypothese 1a). Auch sollte das Aufgabenmerkmal Definitionsgrad (im Sinne schwach definierter Aufgaben) einen negativen Zusammenhang mit der Lösungswahrscheinlichkeit aufweisen (Hypothese 1b). Weiterhin ist ein positiver Zusammenhang des Prozessmerkmals Anzahl von Interaktionen mit der Lösungswahrscheinlichkeit zu erwarten (Hypothese 2a), der sich aber bei einer hohen Anzahl von Interaktionen ins Negative verkehrt (Hypothese 2b). Zudem ist anzunehmen, dass bei höherer Ausprägung des jeweiligen Aufgabenmerkmals (also bei Aufgaben mit einer hohen Zahl benötigter Navigationsschritte und bei schwach definierten Aufgaben) ein stärker positiver Zusammenhang zwischen der Anzahl von Interaktionen und der Lösungswahrscheinlichkeit resultiert (Hypothesen 3a und 3b). Besonders deutlich schließlich sollte der Zusammenhang zwischen der Anzahl von Interaktionen und der Lösungswahrscheinlichkeit bei Aufgaben sein, die hohe Zahlen von Navigationsschritten erfordern *und* schwach definiert sind. Der Unterschied also zwischen dem Zusammenhang von Lösungswahrscheinlichkeit und der Anzahl von Interaktionen zwischen Aufgaben mit mehr und mit weniger benötigten Navigationsschritten sollte dann besonders deutlich zutage treten, wenn die Aufgabe zusätzlich schwach definiert ist. Dies impliziert die Vorhersage

einer Zweifachwechselwirkung zwischen dem Prozessmerkmal Anzahl von Interaktionen und den beiden Aufgabenmerkmalen (Hypothese 4).

## 2 Methode

### 2.1 Stichprobe

Die Grundlage dieser Arbeit bilden Daten aus dem computerbasierten Problemlösetest des PIAAC-Feldtests (OECD, 2013b) aus Deutschland ( $N=661$ ) und Kanada ( $N=411$ ). Die Daten wurden mit einem Multimatrixdesign erhoben, so dass unsere Stichprobe eine Teilstichprobe der PIAAC-Feldteststichprobe darstellt. Die Stichprobe des Feldtests in Deutschland war eine geschichtete Stichprobe aus fünf Bundesländern, die 30% der bundesdeutschen Bevölkerung repräsentieren. Aus diesen wurde eine stratifizierte Zufallsstichprobe von Gemeinden gezogen. Innerhalb dieser Gemeinden wurden Personen zufällig ausgewählt, deren Geburtsdatum zwischen dem 1. Juli 1944 und dem 1. April 1994 lag. In Kanada wurden Personen befragt, die kurz zuvor aus dem kanadischen *labor force survey* (Statistics Canada, 2012) ausrotiert worden waren. Für Analysen wie die hier vorgestellten werden Protokolldateien benötigt, aus denen sich die Interaktion zwischen Probandin und Stimulus rekonstruieren lässt. Diese Daten werden in PIAAC nicht veröffentlicht und liegen uns nur für Deutschland und aufgrund eines bilateralen Kooperationsabkommens für Kanada vor. Dies gibt uns die Möglichkeit, die anhand der deutschen Stichprobe ermittelten Ergebnisse anhand einer zweiten Stichprobe zu replizieren. In Bezug auf die von uns analysierten Effekte gibt es keine Gründe, systematische Unterschiede zwischen deutschen und kanadischen Probandinnen zu erwarten. Häufig zeigen sich Zusammenhänge zwischen Prozessmaßen und Bearbeitungserfolg stabil über Staaten hinweg, auch dann wenn sich in Kompetenz- und Prozessmaßen (parallele) Niveaunterschiede finden (z.B. OECD, 2011, Kapitel 3; Wüstenberg, Greiff, Molnár &

Funke, 2013). In Deutschland waren die Teilnehmerinnen zwischen 16 und 65 Jahren alt ( $M=38$  Jahre;  $SD=13.1$ ). Die kanadischen Teilnehmerinnen waren ebenfalls zwischen 16 und 65 Jahren alt. Da uns für die kanadische Stichprobe aus datenschutzrechtlichen Gründen lediglich Alterskategorien vorliegen, können wir hier keinen Mittelwert und keine Standardabweichung berichten. In der Altersgruppe 16-24 Jahre befanden sich 63 Personen (15.5%), in der Altersgruppe 25-34 Jahre 77 Personen (18.3%), in der Altersgruppe 35-44 Jahre 91 Personen (22.5%), in der Altersgruppe 45-54 Jahre 81 (20.0%) und in der Altersgruppe 55-65 Jahre 96 Personen (23.7%). In der deutschen Stichprobe waren 54.7% der Teilnehmenden männlich und 45.3% weiblich, in der kanadischen Stichprobe 42.2% männlich und 57.8% weiblich. Etwa die Hälfte der Befragten hatten eine Berufsausbildung (Deutschland 48.5%; Kanada: 56.4%) und knapp die Hälfte bzw. ein knappes Viertel der Testteilnehmerinnen einen akademischen Grad (Deutschland: 44.4%; Kanada: 23.5%). Die hier analysierten Daten sind nicht bevölkerungsrepräsentativ.

## **2.2 Materialien**

Von den insgesamt 22 im Feldtest verwendeten TbP-Aufgaben konnten zwei aufgrund von Übersetzungsproblemen nicht gewertet werden. TbP-Aufgaben stellen umfangreiche Interaktionsanforderungen mit simulierten Softwareanwendungen. In einer der veröffentlichten Beispielaufgaben (vgl. [http://www.oecd.org/berlin/PIAAC\\_BspAufgabenDeutschland.pdf](http://www.oecd.org/berlin/PIAAC_BspAufgabenDeutschland.pdf), Datum des Zugriffs: 14.04.2014) z. B. wurden die Testteilnehmerinnen aufgefordert, nach zwei Kriterien eine Internetseite zur Vermittlung von Arbeitsplätzen aus den Suchergebnissen auszuwählen und mit einem Lesezeichen zu versehen. Dabei verweisen die Links in den Suchergebnissen auf Seiten, von denen aus zum Teil weitere Seiten angesteuert werden müssen. Wie viele Seiten insgesamt auszuwählen

sind, geht aus der Aufgabenstellung nicht hervor, ebenso, an welcher Stelle jeweils sich die relevanten Informationen befinden.

### **2.3 Durchführung**

Die Daten wurden in einer Haushaltsbefragung erhoben. Die Probandinnen in Deutschland erhielten eine Gratifikation von maximal € 50.- für eine komplettierte Befragung. In Kanada wurde keine Gratifikation gezahlt. Zunächst wurden durch den Interviewer soziodemographische Daten erfragt. Danach wurden den Teilnehmerinnen zunächst Fragen zur Computervertrautheit sowie ein Kurztest zur Erfassung basaler, für die Bearbeitung des Tests benötigter Computerfertigkeiten vorgelegt. Hierzu zählte z. B. die Bedienung der Computermaus- und Tastatur. Testteilnehmerinnen, deren Computerfertigkeiten sich als für die Testdurchführung unzureichend erwiesen hatten, wurde eine papierbasierte Testvariante vorgelegt. Dies betraf 11.4% der deutschen Stichprobe und 15.9% der kanadischen Stichprobe.

Die TbP-Aufgaben waren auf fünf Booklets verteilt und jede Testteilnehmerin bearbeitete zwei TbP-Booklets. Jede Kombination von TbP-Booklets wurde gemeinsam mit weiteren Booklets vorgegeben, die Aufgaben zu Alltagsmathematischer Kompetenz und Lesekompetenz enthielten. Jedes TbP-Booklet umfasste vier bis sechs Aufgaben, so dass eine Person zwischen acht und zwölf TbP-Aufgaben bearbeiten musste. Insgesamt waren maximal 2 Stunden für die gesamte Testung vorgesehen.

### **2.4 Erhobene Variablen**

Die Aufgabenmerkmale Definitionsgrad und Zahl benötigter Navigationsschritte sowie das Prozessmerkmal Anzahl von Interaktionen wurden wie folgt erfasst:

Die *Zahl benötigter Navigationsschritte* wurde anhand der notwendigen Navigationsschritte (Wechsel zwischen Seiten des Stimulus) bestimmt. Gezählt wurde für die Zahl benötigter Navigationsschritte also jede für die Lösung der Aufgabe benötigte Interaktion mit dem Stimulus, die zur Folge hatte, dass sich der Bildschirminhalt änderte. Dies konnte sowohl ein Wechsel zwischen Teilen einer Software-Anwendung sein (etwa der Wechsel zwischen unterschiedlichen Hypertextseiten) als auch der Wechsel zwischen verschiedenen Softwareanwendungen (etwa der Wechsel zwischen einem Textdokument und einer Tabellenkalkulation). Durchschnittlich waren 4.05 Navigationsschritte nötig, um die Aufgaben zu lösen ( $SD = 2.91$ ). Die oben beschriebene Aufgabe erfordert insgesamt 10 Navigationsschritte. Die Zahl benötigter Navigationsschritte wurde post hoc durch zwei Personen aus dem Kreis der Autoren bestimmt. Die Interrater-Übereinstimmung (Pearson's  $r$ ) dieser Variable betrug .99.

*Definitionsgrad* bezeichnet die Präzision der Aufgabenstellung, d. h. ob alle notwendigen Teilschritte in der Aufgabenstellung benannt wurden oder ob Teile der Aufgabenbearbeitung unklar oder offen gelassen wurden. 55% der hier betrachteten Aufgaben konnten als schwach definiert beschrieben werden. Die oben beschriebene Beispielaufgabe ist in diesem Sinne schwach definiert. Wären – hypothetisch – Informationen darüber in der Aufgabenstellung enthalten, wie viele Links genau auszuwählen sind und wie viele Seiten und Unterseiten anzusteuern sind, sowie nach welchen Phrasen im Einzelnen zu suchen ist, würde es sich um eine im hiesigen Sinne „wohl definierte“ Aufgabe handeln. Da Aufgaben mit mehr benötigten Navigationsschritten häufiger Schritte enthalten, die in der Aufgabenstellung nicht expliziert sind, hängen die beiden Aufgabenmerkmale zusammen: Definitionsgrad (im Sinne schwach definierter Aufgaben) und die Zahl benötigter Navigationsschritte korrelieren zu .65. Der Definitionsgrad der Aufgaben wurde post hoc durch dieselben zwei Personen aus dem Kreis

der Autoren bestimmt wie die Zahl benötigter Navigationsschritte. Die Interrater-Übereinstimmung (Cohen's  $\kappa$ ) betrug .90.

Das Prozessmerkmal *Anzahl der Interaktionen* entspricht der Anzahl von Tastendrücken (Klicks) pro Person und Aufgabe. Eingaben in Textfelder wurden als eine Interaktion gewertet. Damit ist die Anzahl der Interaktionen konzeptuell von der Zahl benötigter Navigationsschritte zu unterscheiden. Erstens wurden wiederholte Interaktionen wiederholt gezählt. Zweitens wurden auch solche Interaktionen gezählt, die kein Navigationsschritt in dem Sinne sind, dass zwischen unterschiedlichen Teilen der Informationsumgebung gewechselt wird. Da die Anzahl der Interaktionen maschinell ermittelt wurde, gehen wir von einer Reliabilität von 1 aus.

Deskriptive Statistiken aller unabhängigen Variablen für beide Stichproben sind in Tabelle 1 enthalten. Für Variablen, die innerhalb von Personen *und* innerhalb von Aufgaben variieren (z.B. Aufgabenscore oder die Anzahl von Interaktionen), musste hier die Beobachtungseinheit die einzelne Reaktion pro Person und Aufgabe sein.

## **2.5 Verteilungen und Datenvorbehandlung**

Aufgrund der über Items und Personen extrem rechtsschiefen Verteilung des Prozessmerkmals Anzahl von Interaktionen (Deutsche Stichprobe: Schiefe  $v = 3.37$ , Kanadische Stichprobe:  $v = 4.56$ ) wurde die Anzahl von Interaktionen logarithmiert. Durch die Logarithmierung ergab sich eine leicht rechtsschiefe Verteilung, der Betrag der Schiefe wurde deutlich reduziert (deutsche Stichprobe:  $v = -0.35$ , kanadische Stichprobe:  $v = -0.21$ ). Um eine bessere Interpretation der logarithmischen Skala zu gewährleisten, wurde die logarithmierte Anzahl von Interaktionen zusätzlich über Items und Personen z-standardisiert. Das Aufgabenmerkmal Zahl benötigter Navigationsschritte wurde z-standardisiert, das dichotome Aufgabenmerkmal Definitionsgrad wurde zentriert. Die Zentrierung diente in

beiden Fällen dem Zweck, den Haupteffekt jedes Prädiktors bei mittlerer Ausprägung aller anderen Prädiktoren schätzen zu können (vgl. Aiken & West, 1991).

## 2.6 Auswertungsstrategie

Zur Prüfung der Hypothesen wurden generalisierte gemischte lineare Modelle (GLMMs) verwendet (vgl. Baayen, Davison, & Bates, 2008; De Boeck et al., 2011). In einem solchen Modell werden die Logit-transformierten Wahrscheinlichkeiten (log-odds), jeweils eine bestimmte Aufgabe zu lösen durch feste Effekte und Zufallseffekte erklärt. Feste Effekte sind konstant über Einheiten einer Gruppe, Zufallseffekte können dagegen über diese Einheiten (z. B. Personen, Schulklassen) variieren. Die Einflüsse von Personen („Problemlösekompetenz“) und Aufgaben („Leichtigkeit“) auf die Lösungswahrscheinlichkeit wurden als Zufallseffekte spezifiziert. Die Effekte des Prozessmerkmals Anzahl von Interaktionen und der entsprechende quadratische Term, sowie die Effekte der Aufgabenmerkmale und ihre Wechselwirkungen wurden als feste Effekte in das Modell eingefügt. Das folgende Modell wurde geschätzt:

$$\begin{aligned} \ln[p/(1-p)] = & \beta_0 + \beta_1(\text{Benötigte Navigationsschritte}) + \beta_2(\text{Definitionsgrad}) \\ & + \beta_3(\text{Benötigte Navigationsschritte} \times \text{Definitionsgrad}) + \beta_4(\text{Anzahl Interaktionen}) \\ & + \beta_5(\text{Anzahl Interaktionen})^2 + \beta_6(\text{Anzahl Interaktionen} \times \text{benötigte Navigationsschritte}) \\ & + \beta_7(\text{Anzahl Interaktionen} \times \text{Definitionsgrad}) + \beta_8(\text{Anzahl Interaktionen} \times \text{benötigte} \\ & \text{Navigationsschritte} \times \text{Definitionsgrad}) + b_{0p} + b_{0i} \end{aligned} \quad (1)$$

$p$  kennzeichnet die Wahrscheinlichkeit einer erfolgreichen Aufgabenbearbeitung.  $\beta_0$  ist das Intercept,  $\beta_1$  bis  $\beta_8$  sind feste Effekte der Prädiktoren und ihrer Wechselwirkungen.  $b_{0p}$  ist die als Zufallseffekt modellierte personenspezifische Adjustierung des Intercepts (entspricht der Problemlösekompetenz einer Person).  $b_{0i}$  ist die als Zufallseffekt modellierte itemspezifische Adjustierung des Intercepts (entspricht der Leichtigkeit eines Items). Zur

Schätzung der Modelle wurde die R-Umgebung (R Core Team, 2012) in Verbindung mit dem Paket lme4 (Bates, Maechler, & Bolker, 2012) genutzt. Für beide Stichproben wurde der Effekt aller Prädiktoren simultan geschätzt. Die Verteilung der Problemlösekompetenz war eingipflig, wobei sich in der kanadischen Stichprobe eine etwas stärker rechtsschiefe Verteilung zeigte, als qua Normalverteilungsannahme zu vermuten. Das GLMM ist jedoch gegenüber leichten Verletzungen der Normalverteilungsannahme robust (McCulloch & Neuhaus, 2011). Als Effektgröße berichten wir den Anteil der durch die festen, sowie die durch die festen und die zufälligen Effekte erklärten Anteile erklärter Varianz für GLMMs in Anlehnung an Nakagawa und Schielzeth (2013).

## **2.7 Fehlende Werte und N pro Aufgabe**

Abgesehen von den durch das Multimatrix-Design bedingt fehlenden Werten kamen in einzelnen Aufgaben fehlende Werte durch Nicht-Bearbeitung einzelner Aufgaben durch einzelne Personen zustande. Dies betraf 107 (2.9%) der zur Verfügung stehenden Beobachtungen in der deutschen und 157 (6.6%) in der kanadischen Stichprobe. Pro Aufgabe standen in der deutschen Stichprobe zwischen 140 und 194 und in der kanadischen Stichprobe zwischen 95 und 127 Beobachtungen zur Verfügung.

## **3 Ergebnisse**

Deskriptive Statistiken und Korrelationen aller beteiligten Variablen sind in Tabelle 1 enthalten. Das  $\alpha$ -Fehlerniveau wurde auf .05 festgesetzt. Wir berichten für die festen Effekte unstandardisierte Regressionskoeffizienten. Die metrischen Prädiktoren sind z-standardisiert. Die Regressionskoeffizienten lassen sich daher als diejenige Änderung in den log-odds einer erfolgreichen Aufgabebearbeitung interpretieren, die mit einer Steigerung des metrischen Prädiktors um eine Standardabweichung einhergeht. Das dichotome Merkmal

Definitionsgrad der Aufgabe wurde als gewichtete Dummy-Variable mit einem Mittelwert von Null und einem Abstand von eins zwischen den beiden Kategorien eingeführt. Effekte dieses Aufgabenmerkmals lassen sich daher als vorhergesagte Änderung der log-odds beim Übergang von wohl zu schwach definierten Aufgaben interpretieren. Wir berichten zunächst die Ergebnisse für die deutsche Stichprobe und gehen dann auf Gemeinsamkeiten und Unterschiede zur kanadischen Stichprobe ein.

### **3.1 Ergebnisse für die deutsche Stichprobe**

#### **3.1.1 Zufallseffekte**

Die personenspezifische Varianzkomponente wurde als 0.68 geschätzt. Ein Log-likelihood-Differenztest gegen ein Modell, in dem diese Varianzkomponente zu Null restringiert war, ergab eine signifikante Modellverschlechterung,  $\chi^2(1) = 67.59, p < .05$ . Dies war zu erwarten angesichts der Tatsache, dass diese Varianzkomponente zufällige Unterschiede zwischen Personen (d.h. ihre Problemlösekompetenz) abbildet. Die aufgabenspezifische Varianzkomponente wurde als 1.32 geschätzt. Ein Log-likelihood-Differenztest gegen ein Modell, in dem diese Varianzkomponente zu Null restringiert war, ergab eine signifikante Modellverschlechterung,  $\chi^2(1) = 521.50, p < .05$ . Dies war wiederum zu erwarten, da diese Varianzkomponente zufällige Unterschiede zwischen Aufgaben (d.h. ihre Leichtigkeit) abbildet.

#### **3.1.2 Feste Effekte**

Die festen Effekte für die deutsche Stichprobe sind in Tabelle 2 (Spalten 2-5) zusammengefasst.

##### **3.1.2.1 Effekte der Aufgabenmerkmale**

Wie in Hypothese 1a angenommen, hatte das Aufgabenmerkmal „Benötigte Navigationsschritte“ einen negativen Zusammenhang mit der Lösungswahrscheinlichkeit. Der Anteil aufgeklärter Varianz war allerdings mit unter einem Prozent gering. Für Aufgaben mit einer niedrigen Zahl benötigter Navigationsschritte (eine Standardabweichung unter dem Mittelwert) wurde im Mittel eine Lösungswahrscheinlichkeit von .53 vorhergesagt. Für Aufgaben mit einer hohen Zahl benötigter Navigationsschritte dagegen betrug die vorhergesagte Lösungswahrscheinlichkeit nur .14. Für das Aufgabenmerkmal Definitionsgrad fand sich anders als in Hypothese 1b angenommen kein signifikanter Zusammenhang mit der Lösungswahrscheinlichkeit.

### **3.1.2.2 Effekte des Prozessmerkmals Anzahl von Interaktionen**

Wie in Hypothese 2a vermutet, stieg die Lösungswahrscheinlichkeit mit der Anzahl von Interaktionen (Tabelle 2, Zeile 5). Der mit diesem Effekt assoziierte Anteil aufgeklärter Varianz von 14% entsprach einem großen Effekt. Dabei war der Zusammenhang in Übereinstimmung mit Hypothese 2b nicht linear. Wie an dem signifikanten negativen Regressionskoeffizienten für den quadratischen Trend abgelesen werden kann, stieg die Lösungswahrscheinlichkeit zunächst mit zunehmender Anzahl von Interaktionen an, fiel dann aber nach Erreichen eines Maximums wieder ab. Im Mittel über alle Aufgaben wurde die maximale Lösungswahrscheinlichkeit bei einer Anzahl von Interaktionen erreicht, die 1.65 Standardabweichungen über dem Mittelwert liegt (entsprechend 20 Interaktionen in der ursprünglichen Metrik der Variable), und betrug dann .56. Im Kontrast dazu wurde bei einer niedrigen Anzahl von Interaktionen (-1 *SD*, entsprechend 8 Interaktionen) eine Lösungswahrscheinlichkeit von nur .05 erreicht. Der durch den quadratischen Trend aufgeklärte Varianzanteil von 9% entsprach einem großen Effekt.

### **3.1.2.3 Wechselwirkung zwischen der Anzahl von Interaktionen und der Anzahl benötigter Navigationsschritte**

Der Effekt des Prozessmerkmals Anzahl Interaktionen variierte in Übereinstimmung mit Hypothese 3a deutlich in Abhängigkeit von der Anzahl benötigter Navigationsschritte. Der mit der entsprechenden Wechselwirkung assoziierte Anteil erklärter Varianz (12%) entsprach einem großen Effekt. Wie vermutet, war der Effekt der Anzahl von Interaktionen vor allem dann prononciert, wenn die Aufgaben eine hohe Zahl von Navigationsschritten (+ 1SD) verlangte (s. die mit Dreiecken gekennzeichnete durchgezogene Linie in Abbildung 1a sowie die mit Dreiecken gekennzeichneten Linien in Abbildung 3c). In diesem Fall ergab sich ein einfacher Regressionskoeffizient von  $\beta = 3.42$  ( $SE = 0.35$ ),  $z = 9.76$ ,  $p < .05$ . Verlangte die Aufgabe nur eine niedrige Zahl von Navigationsschritten, konnte kein Zusammenhang zwischen der Anzahl von Interaktionen und dem Lösungserfolg gefunden werden,  $\beta = -0.03$  ( $SE = 0.11$ ),  $z = 0.27$ ,  $p = .61$  (s. die mit Dreiecken gekennzeichnete gestrichelte Linie in Abbildung 1a).

### **3.1.2.4 Wechselwirkung zwischen der Anzahl von Interaktionen und dem Definitionsgrad der Aufgabe**

In Hypothese 3b war vermutet worden, dass vor allem bei schwach definierten Aufgaben die Anzahl von Interaktionen stark positiv mit der Lösungswahrscheinlichkeit kovariert. Dies war nicht der Fall. Im Gegenteil waren wohl definierte Aufgaben diejenigen, bei denen ein stärkerer (positiver) Zusammenhang zwischen der Anzahl von Interaktionen und der Lösungswahrscheinlichkeit resultierte (siehe Tabelle 2, Zeile 8). Der mit dieser Interaktion assoziierte Anteil aufgeklärter Varianz (2%) entsprach einem kleinen Effekt. Unter Heranziehung einfacher Regressionskoeffizienten kann die Wechselwirkung wie folgt illustriert werden: Für schwach definierte Aufgaben ergab sich ein deutlicher

Zusammenhang zwischen der Anzahl von Interaktionen und der Lösungswahrscheinlichkeit,  $\beta = 0.81$  ( $SE = 0.11$ ),  $z = 7.43$ ,  $p < .05$ , was für sich genommen der Erwartung entsprach (s. die mit Dreiecken gekennzeichnete durchgezogene Linie in Abbildung 1b). Allerdings wurde dieser Zusammenhang erwartungswidrig bei wohl definierten Aufgaben nicht etwa abgeschwächt, sondern fiel noch einmal stärker aus,  $\beta = 2.58$  ( $SE = 0.32$ ),  $z = 7.99$ ,  $p < .05$  (s. die mit Dreiecken gekennzeichnete gestrichelte Linie für Deutschland in Abbildung 1b).

### **3.1.2.5 Zweifachwechselwirkung zwischen der Anzahl von Interaktionen, benötigten**

*Navigationsschritten und dem Definitionsgrad der Aufgabe.* Ebenfalls entgegen den (in Hypothese 4 formulierten) Erwartungen fiel die Zweifachwechselwirkung zwischen dem Prozessmerkmal und den beiden Aufgabenmerkmalen negativ aus (s. Tabelle 2, Zeile 9). Der mit der Zweifachwechselwirkung assoziierte Anteil erklärter Varianz (1%) fiel gering aus. Unter Heranziehung einfacher (bedingter) Wechselwirkungen kann die Zweifachwechselwirkung wie folgt illustriert werden: Bei schwach definierten Aufgaben fand sich eine Wechselwirkung zwischen der Anzahl von Interaktionen und der Anzahl benötigter Navigationsschritte von  $\beta = 0.57$  ( $SE = 0.12$ ),  $z = 4.65$ ,  $p < .05$ . Diese Wechselwirkung fiel bei wohl definierten Aufgaben nicht schwächer, sondern stärker aus,  $\beta = 2.87$  ( $SE = 0.38$ ),  $z = 7.99$ ,  $p < .05$ .

## **3.2 Gemeinsamkeiten und Unterschiede der deutschen und der kanadischen Stichprobe**

### **3.2.1 Zufallseffekte**

Die personen- und aufgabenspezifischen Zufallseffekte fielen in der kanadischen Stichprobe ähnlich aus wie in der deutschen, erneut ergab sich ein signifikanter personenspezifischer Zufallseffekt von  $0.83$ ,  $\chi^2(1) = 1528.60$ ,  $p < .05$ , sowie ein signifikanter aufgabenspezifischer Zufallseffekt von  $1.19$ ,  $\chi^2(1) = 1982.50$ ,  $p < .05$ .

### 3.2.2 Feste Effekte

Die festen Effekte in der kanadischen Stichprobe replizierten die entsprechenden Effekte der deutschen Stichprobe fast vollständig (vgl. die Spalten 2 und 6 in Tabelle 2, vgl. auch Abbildung 1c mit Abbildung 1d). Das Muster von positiven und negativen, sowie von signifikanten und insignifikanten Effekten war vollständig identisch, und auch die Parameterschätzungen selbst ähnelten sich stark, ebenso wie die mit den einzelnen festen Effekten assoziierten Anteile erklärter Varianz. Der einzige auffällige Unterschied lag in einem der einfachen Regressionskoeffizienten. Bei Aufgaben, die eine geringe Anzahl von Navigationsschritten erfordern, hatte sich in der deutschen Stichprobe kein Zusammenhang zwischen der Anzahl von Interaktionen und der Lösungswahrscheinlichkeit gefunden. In der kanadischen Stichprobe dagegen blieb dieser Zusammenhang auch bei Aufgaben erhalten, die nur eine geringe Zahl von Navigationsschritten erfordern ( $-1SD$ ),  $\beta = 0.45$  ( $SE = 0.15$ ),  $z = 3.03$ ,  $p < .05$  (vgl. die mit Kreisen gekennzeichnete gestrichelte Linie in Abbildung 1a). Für Aufgaben, die eine hohe Zahl von Navigationsschritten erfordern, war auch in der kanadischen Stichprobe ein deutlich stärkerer Zusammenhang mit der Lösungswahrscheinlichkeit zu beobachten,  $\beta = 3.06$  ( $SE = 0.42$ ),  $z = 7.34$ ,  $p < .05$ , der ähnlich ausfiel wie in der deutschen Stichprobe (vgl. die mit Punkten gekennzeichnete durchgezogene Linie in Abbildung 1a).

## 4 Diskussion

Wir haben in dieser Arbeit gezeigt, wie sich beim technologiebasierten Problemlösen der Problemlöseerfolg aus der Interaktion mit dem technologischen System vorhersagen lässt. In Übereinstimmung mit unseren Annahmen ist die Anzahl von Interaktionen ein guter Prädiktor für den Problemlöseerfolg, wobei eine maximale Lösungswahrscheinlichkeit

gemittelt über alle Aufgaben bei einer Anzahl von Interaktionen erreicht wird, die ca. um zwei Standardabweichungen über dem Mittel liegt. Dieses Resultat ist konsistent mit Ergebnissen zum informationsbezogenen technologiebasierten Problemlösen, und erweitert diese Befundlage auf Probleme, bei denen das Ziel nicht in Informationssuche, Textverstehen oder Wissenserwerb, sondern z.B. in der Organisation von Tabellenblättern besteht. Moderiert wurde der Zusammenhang dabei erwartungsgemäß durch die Zahl der benötigten Navigationsschritte: In TbP-Aufgaben, die eine hohe Zahl von Navigationsschritten erfordern, ist die Anzahl von Interaktionen ein sehr starker Prädiktor für den Problemlöseerfolg. Bei solchen Aufgaben ist die Lösungswahrscheinlichkeit nahe 0, wenn eine unterdurchschnittliche Anzahl von Interaktionen ausgeführt wird, wohingegen die Lösungswahrscheinlichkeit dann, wenn ausreichend mit der Aufgabe interagiert wird, 1 approximiert.

Entgegen den Erwartungen fielen die Wechselwirkung zwischen der Anzahl von Interaktionen und dem Definitionsgrad der TbP-Aufgaben sowie die Zweifachwechselwirkung zwischen Anzahl von Interaktionen, Definitionsgrad und Zahl benötigter Navigationsschritte aus. Abbildung 1c-d hilft zu verstehen, warum dies so war. Bei wohl definierten Aufgaben, die eine hohe Zahl von Navigationsschritten erfordern (die mit Dreiecken gekennzeichneten durchgezogenen Kurven in den Abbildungen 1c und 1d), war die Lösungswahrscheinlichkeit bei unterdurchschnittlicher Anzahl von Interaktionen praktisch 0 und näherte sich bereits bei einer Anzahl von Interaktionen von einer Standardabweichung über dem Mittelwert 1. Bei schwach definierten Aufgaben, die eine hohe Zahl von Navigationsschritten erfordern (die mit Dreiecken gekennzeichneten gestrichelten Kurven in den Abbildungen 1c und 1d) dagegen war die Lösungswahrscheinlichkeit bei unterdurchschnittlicher Anzahl von Interaktionen ebenfalls nahe 0 und stieg mit zunehmender Anzahl von Interaktionen dann vergleichsweise langsam.

Wie aus den Abbildungen 1c und 1d hervorgeht, hatte dies auch zur Folge, dass – immer eine hohe Zahl benötigter Navigationsschritte vorausgesetzt – die Lösungswahrscheinlichkeit bei wohl definierten Aufgaben über der Lösungswahrscheinlichkeit schwach definierter Aufgaben lag, wenn eine hohe Anzahl von Interaktionen vorliegt. Anders ausgedrückt: Bei wohl definierten Aufgaben mit klarem Lösungsweg führte eine hohe Anzahl von Interaktionen zu Lösungswahrscheinlichkeiten nahe 1. Bei schwach definierten Aufgaben dagegen führte auch eine hohe Anzahl von Interaktionen zu Lösungswahrscheinlichkeiten von lediglich ca. 80%.

Insgesamt stützen die vorliegenden Analysen und Befunde die Annahme, dass es eher ein zu zurückhaltender und zu wenig explorativer Umgang mit Computertechnologie ist, der Personen am erfolgreichen Lösen technologiebasierter Probleme hindert, als ein zur forscher Umgang. Will man diesen Befund kausal interpretieren (was durch geeignete Experimente abzusichern wäre), sollten Instruktionsdesigns für den Umgang mit technologiebasierten Problemen einen solchen Umgang fördern. Wie ein solcher Umgang mit der Vorsicht zu integrieren ist, die bei technologiebasierten Problemen auch angezeigt sein kann (i. e. *nicht* auf einen Link in einer E-Mail klicken, in der „Sparkasse, Internet-Banking-Abteilung“ nach „Ihrer Angaben“ zu „Um zu unserem neuen Sicherheits-System zu verbinden und zum Schutz vor Betrug“ fragt): Dies zu überlegen ist eine der Herausforderungen für das Design von Curricula, die Schülerinnen auf das Lösen technologiebasierter Probleme vorbereiten.

Die Analyseeinheiten in unserer Arbeit mussten individuelle *Aufgabenbearbeitungsprozesse* sein. Anders hätten Interaktionen zwischen Prozess- und Aufgabenmerkmalen nicht modelliert werden können. Dies bedeutet aber natürlich auch, dass wir bis dato nichts darüber wissen, inwieweit *Personen* im Hinblick auf adäquates Aufgabenbearbeitungsverhalten konsistent sind. Die Feststellung, dass die Bearbeitung einer

individuellen Aufgabe seitens einer individuellen Person in der in dieser Arbeit gezeigten Art und Weise mit der Wahrscheinlichkeit zusammenhängt, dass diese Person die in Frage stehende Aufgabe löst, wird davon allerdings nicht beeinträchtigt. Ob also die Art und Weise der Bearbeitung von TbP-Aufgaben eher ein *state* oder ein *trait* ist, muss künftige Forschung klären. Hiervon kann dann auch abhängig gemacht werden, wo einschlägige Interventionen vor allem ansetzen sollten – ob eher auf der Seite der Instruktion für individuelle Aufgaben, oder eher auf der Seite von Instruktionsprogrammen, mit der personenseitig der Zugriff auf technologiebasierte Probleme beeinflusst werden soll.

Unter einer methodologischen Perspektive machen diese Erwägungen auch die Limitierungen deutlich, mit denen der hier verfolgte Ansatz behaftet ist, substanzwissenschaftliche Fragestellungen anhand von Daten aus Large-Scale-Assessments zu verfolgen. Einerseits sind bei einem solchen Untersuchungsansatz zwar größere und heterogenere Stichproben zugänglich, als sie in Untersuchungen im psychologischen Labor betrachtet werden können, was zweifellos der Validität und Generalisierbarkeit der Ergebnisse zugute kommt. Andererseits jedoch besteht hinsichtlich der analysierten Daten häufig großer Interpretationsspielraum. Dies gilt auch dann, wenn nicht nur Informationen über Ergebnisse, sondern auch über Prozesse vorliegen. Exemplarisch kann man das an dem hier betrachteten Indikator „Anzahl von Interaktionen“ sehen: Dem Indikator kommen für unterschiedliche Abschnitte seiner Verteilung unterschiedliche Bedeutungen zu. Im unteren Bereich der Verteilung kennzeichnet der Indikator vermutlich aufgabenadäquates Problemlöseverhalten, im oberen Bereich der Verteilung Desorientierung. Auch beschreibt jeder einzelne Prozessindikator nur Teile des Problemlöseprozesses. Zu ergänzen wären also Analysen mit stärkerem Bezug zur Problemdomäne „Technologiebasiertes Problemlösen“. Beispielsweise ist aus dem Bereich der Informationssuche im Internet die semantische Kohärenz zwischen Suchziel und ausgewähltem Hyperlink ein zentraler Bestandteil

erfolgreicher Problemlösung (Klein & Kominski, 2013). Für derartige Untersuchungen wäre allerdings speziell geeignetes Versuchsmaterial zu konstruieren, da sich solche spezifischen Indices i. d. R. nur für die Beschreibung von Prozessen in eng definierten Domänen eignen. Eine zweite Limitierung der Generalisierbarkeit von Ergebnissen wie den vorliegenden ergibt sich aus der Tatsache, dass ausschließlich Personen mit hinreichenden Computerfähigkeiten untersucht werden konnten. Und eine ganz allgemeine Beschränkung der hier vorgestellten Ergebnisse ergibt sich schließlich aus der korrelativen Natur unserer Daten. In den vorliegenden Analysen ist zwar der Effekt jedes der beiden betrachteten Itemmerkmale durch das jeweils andere Itemmerkmal kontrolliert. Bereits ein Blick in die PIAAC-Rahmenkonzeption für technologiebasiertes Problemlösen informiert allerdings darüber, dass die PIAAC TbP-Aufgaben noch in einer großen Zahl von weiteren Merkmalen variieren. Abbilden lassen sich diese Merkmale jedoch nicht, so lange nicht eine größere Zahl von Aufgaben vorliegt. Dies zu realisieren wiederum ist schwierig in einem Feld, wo die Aufgaben verglichen mit sonstigen in Large-Scale-Assessments zum Einsatz kommenden Aufgaben in der Administrierung zeitaufwändig sind. Forschungsstrategisch ist also auch vor diesem Hintergrund der hier verfolgte Ansatz der Analyse großer Datensätze zu verbinden mit experimentellen Untersuchungen, in denen die interessierenden Aufgabenmerkmale besser zu kontrollieren sind, und die auch eine Triangulation durch weitere Prozessmaße, wie Blickbewegungen oder Protokolle lauten Denkens, ermöglichen würden.

### **Autorenhinweis**

Die vorliegende Arbeit ist im Rahmen des von der Deutschen Forschungsgemeinschaft (DFG) geförderten Projekts „Prozessdaten zur Kompetenzmodellierung“ (Leitung: Frank Goldhammer, Johannes Naumann und Heiko Rölke, GO 1979/1-1) entstanden. Wir danken der

DFG für ihre Unterstützung. Wir danken außerdem Beatrice Rammstedt und der PIAAC-Arbeitsgruppe bei GESIS, Mannheim (<http://www.gesis.org>), und dem Bundesministerium für Bildung und Forschung für den Zugang zu den Daten des deutschen PIAAC-Feldtests. Wir danken Sylvie Grenier von Statistics Canada (<http://www.statcan.gc.ca/>) für den Zugang zu den Daten des kanadischen PIAAC-Feldtests. Johannes Naumann ist jetzt an der Goethe-Universität Frankfurt. E-Mail: [j.naumann@em.uni-frankfurt.de](mailto:j.naumann@em.uni-frankfurt.de).

## Literatur

- Aiken, L. S. & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.
- Baayen, R.H., Davidson, D.J. and Bates, D.M. (2008) Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59, 390-412.
- Bates, D., Maechler, M., & Bolker, B. (2012). *lme4: Linear mixed-effects models using Eigen and Eigenfaces*. R package version 0.999999-0. Retrieved January 14, 2014 from <http://CRAN.R-project.org/package=lme4>
- Bell, B.S., & Kozlowski, S.W.J. (2008). Active learning: Effects of core training design elements on self-regulatory processes, learning, and adaptability. *Journal of Applied Psychology*, 93, 296–316.
- Brand-Gruwel, S., Wopereis, I., & Walraven, A. (2009). A descriptive model of Information Problem Solving while using Internet. *Computers & Education*, 53, 1207-1217.
- De Boeck, P., Bakker, M., Zwitser, R., Nivard, M., Hofman, A., Tuerlinckx, F., & Partchev, I. (2011). The estimation of item response models with the lmer function from the lme4 package in R. *Journal of Statistical Software*, 39, 1-28.

- Dormann, T., & Frese, M. (1994). Error management training: Replication and the function of exploratory behavior. *International Journal of Human–Computer Interaction*, 6, 365–372.
- Funke, J. (2006). Komplexes Problemlösen. In J. Funke (Hrsg.), *Denken und Problemlösen (Enzyklopädie der Psychologie, Themenbereich C: Theorie und Forschung, Serie II: Kognition, Band 8)* (S. 375-445). Göttingen: Hogrefe.
- Greiff, S., Kretschmar, A., & Leutner, D. (2014). Editorial zum Themenheft Problemlösen in der Pädagogischen Psychologie. *Zeitschrift für Pädagogische Psychologie*.
- Herder, E., & Juvina, I. (2004). Discovery of individual user navigation styles. In G. D. Magoulas & S. Y. Chen (Eds.), *Adaptive hypermedia AH2004 workshop on individual differences in adaptive hypermedia*. Eindhoven, The Netherlands: Springer.
- Keith, N., & Frese, M. (2008). Effectiveness of error management training: A metaanalysis. *Journal of Applied Psychology*, 93, 59–69.
- 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.
- Klein, F. C. & Kominski, J. F. (2013). Missing links in middle school: Developing use of disciplinary relatedness in evaluating Internet search results. *PLOS ONE*, 8, 1-5.
- McCulloch, C. E. & Neuhaus, J. M. (2011). Misspecifying the shape of a random effects distribution: Why getting it wrong may not matter. *Statistical Science*, 26, 388-402.
- Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining  $R^2$  from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4, 133-142.
- OECD. (2011). *PISA 2009 results vol VI: Student online. Digital technologies and performance*. Paris: OECD Publishing.

- OECD. (2012). *Literacy, numeracy and problem solving in technology-rich environments: Framework for the OECD survey of adult skills*. Paris: OECD Publishing.
- OECD. (2013a). *OECD skills outlook 2013: First results from the survey of adult skills*. Paris: OECD Publishing.
- OECD. (2013b). *Technical report of the survey of adult skills (PIAAC) (Pre-publication copy)*. Paris: OECD. Retrieved January 14, 2014 from [http://www.oecd.org/site/piaac/\\_Technical\\_Report\\_17OCT13.pdf](http://www.oecd.org/site/piaac/_Technical_Report_17OCT13.pdf)
- Öllinger, M. & Knoblich, G. (2006). Lösen einfacher Probleme. In J. Funke & P. A. Frensch (Hrsg.), *Handbuch der Allgemeinen Psychologie – Kognition (Handbuch der Psychologie, Bd. 5)* (S. 431-438). Göttingen: Hogrefe.
- Powell, A. L. (2013). Computer anxiety: Comparison of research from the 1990s and 2000s. *Computers in Human Behavior*, 29, 2337-2381.
- R Core Team (2012). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved January 14, 2014 from <http://www.R-project.org/>
- Rammstedt, B. (2013). PIAAC 2012: Die wichtigsten Ergebnisse im Überblick. In B. Rammstedt (Hrsg.), *Grundlegende Kompetenzen Erwachsener im internationalen Vergleich. Ergebnisse von PIAAC 2012* (S. 11-20). Münster: Waxmann.
- Richter, T., Naumann, J., Brunner, M., & Christmann, U. (2005). Strategische Verarbeitung beim Lernen mit Text und Hypertext. *Zeitschrift für Pädagogische Psychologie*, 19, 5-22.
- Statistics Canada (2012). *Guide to the labor force survey*. Ottawa: Statcan. Retrieved June 17, 2014 from <http://www.statcan.gc.ca/pub/71-543-g/71-543-g2012001-eng.pdf>

Wüstenberg, S., Greiff, S., Molnár, G & Funke, J. (2013). Cross-national gender differences in complex problem solving and their determinants. *Learning and Individual Differences, 29*, 18-29.

Prof. Dr. Johannes Naumann

Quantitative Methoden der Erziehungswissenschaft

Institut für Allgemeine Erziehungswissenschaft

Senckenberanlage 31

60054 Frankfurt am Main

Deutschland

[j.naumann@em.uni-frankfurt.de](mailto:j.naumann@em.uni-frankfurt.de)

Tabelle 1

*Deskriptive Statistiken und Korrelationen*

Deutschland ( $k_{ij} = 3521$ ) <sup>a</sup>					Korrelationen			
	<i>M</i>	<i>SD</i>	<i>Min.</i>	<i>Max.</i>	1.	2.	3.	4.
1. Aufgabe schwach definiert <sup>b</sup>	0.55	0.50	0	1				
2. Zahl benötigter Navigationsschritte <sup>b</sup>	4.05	2.91	1	10	.65 <sup>b,c</sup>			
3. Anzahl Interaktionen	22.54	27.54	1	399	.00 <sup>c</sup>	.15		
4. Anzahl Interaktionen logarithmiert	2.43	1.30	0	5.99	.13 <sup>c</sup>	.32	(.77)	
5. Score <sub>ij</sub> <sup>e</sup>	0.36	0.48	0	1	-.04 <sup>d</sup>	-.09 <sup>c</sup>	.00 <sup>c</sup>	.13 <sup>c</sup>
Kanada ( $k_{ij} = 2212$ ) <sup>a</sup>					Korrelationen			
	<i>M</i>	<i>SD</i>	<i>Min.</i>	<i>Max.</i>	1.	2.	3.	4.
1. Aufgabe schwach definiert <sup>b</sup>	0.55	0.50	0	1				
2. Zahl benötigter Navigationsschritte <sup>b</sup>	4.05	2.91	1	10	.65 <sup>b,c</sup>			
3. Anzahl Interaktionen	20.40	27.72	1	449	-.02 <sup>c</sup>	.15		
4. Anzahl Interaktionen logarithmiert	2.29	1.31	0	6.11	.09 <sup>c</sup>	.30	(.75)	
5. Score <sub>ij</sub> <sup>e</sup>	0.31	0.46	0	1	-.06 <sup>d</sup>	-.08 <sup>c</sup>	.04 <sup>c</sup>	.18 <sup>c</sup>

*Anmerkungen.* Korrelationen zwischen arithmetisch voneinander abhängigen Variablen stehen in Klammern, Korrelationen sind Produkt-Moment Korrelationen soweit nicht anders angegeben.

<sup>a</sup>Anzahl von Antworten über Personen *i* und Aufgaben *j*. <sup>b</sup>Schwach definierte Aufgaben (1) vs. wohldefinierte Aufgaben (0), Mittelwerte, Standardabweichung und Korrelation auf Aufgabenebene ( $k = 20$ ). <sup>c</sup>Punkt-biseriale Korrelation. <sup>d</sup>Phi-Koeffizient. <sup>e</sup>Lösung (1) vs. Nicht-Lösung (0) jeder Person *i* bei jeder von dieser Person bearbeiteten Aufgabe *j*.

Tabelle 2

*Feste Effekte der Aufgabenmerkmale, des Prozessmerkmals und ihrer Interaktionen in der deutschen und kanadischen Stichprobe des PIAAC-Feldtests*

Feste Effekte	Deutsche Stichprobe				Kanadische Stichprobe			
	Effekt $\beta$	(SE)	$z$	$\Delta R^2_m$	Effekt $\beta$	(SE)	$z$	$\Delta R^2_m$
Intercept	-0.81	(0.39)			-1.36	(0.39)		
Benötigte Navigationsschritte	-0.96	(0.44)	-2.19*	<.01	-1.02	(0.43)	-2.35*	.01
Aufgabe schwach definiert	0.46	(0.41)	1.12	<.01	0.40	(0.41)	0.99	<.01
Benötigte Navigationsschritte $\times$ Aufgabe schwach definiert	-0.21	(0.46)	-0.45	.01	-0.04	(0.46)	-0.10	<.01
Anzahl Interaktionen	1.69	(0.16)	10.36*	.14	1.76	(0.20)	8.87*	.18
Anzahl Interaktionen quadriert	-0.51	(0.06)	-8.35*	.09	-0.35	(0.08)	-4.17*	.06
Anzahl Interaktionen $\times$ Benötigte Navigationsschritte	1.72	(0.20)	8.52*	.12	1.30	(0.24)	5.36*	.09
Anzahl Interaktionen $\times$ Aufgabe schwach definiert	-0.88	(0.18)	-4.98	.02	-0.63	(0.21)	-3.04	.02
Anzahl Interaktionen $\times$ Benötigte Navigationsschritte $\times$ Aufgabe schwach definiert	-1.15	(0.20)	-5.80	.02	-0.89	(0.23)	-3.83	.01
Modellgüte	$R^2_m = .29, R^2_c = .56$				$R^2_m = .29, R^2_c = .56$			

*Anmerkungen.*  $R^2_m$ : Marginales  $R^2$ , entspricht dem durch die festen Effekte erklärten Varianzanteil.  $R^2_c$ : Konditionales  $R^2$ , entspricht dem durch die festen und Zufallseffekte erklärten Varianzanteil (s. Nakagawa & Schielzeth, 2013).  $\Delta R^2_m$ : Differenz zwischen  $R^2_m$  für das Modell mit allen Prädiktoren und  $R^2_m$  für ein Modell, aus dem der respektive Prädiktor ausgeschlossen ist.

\*  $p < .05$  (einseitige Tests).

## Abbildung

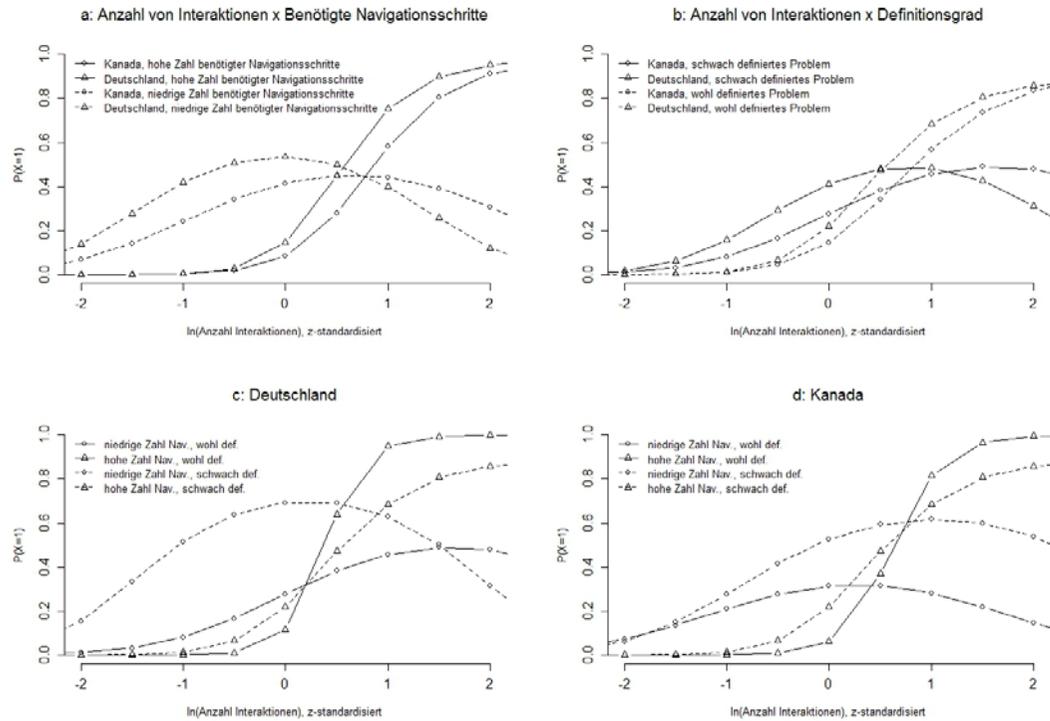


Abbildung 1. Einfachwechselwirkungen zwischen der Anzahl von Interaktionen und Zahl der benötigten Navigationsschritte sowie zwischen der Anzahl von Interaktionen und dem Definitionsgrad der Aufgabe (Abbildungen 1a und 1b) sowie vorhergesagte Lösungswahrscheinlichkeit für Aufgaben mit hoher und niedriger Zahl benötigter Navigationsschritte bei schwach definierten und wohl definierten Aufgaben in der deutschen sowie in der kanadischen Stichprobe (Abbildungen 1c und 1d).