



Entwicklung eines Messinstruments zu den Grunderfahrungen des Informatikunterrichts

ANDREAS KIENER

Zusammenfassung. The three basic experiences of computer science education (GI) take into account the personal perceptions and attitudes of students to computer science education. The aim of this study is to develop an inventory to capture these learners' perceptions and perspectives in order to select content or to track learners' development in relation to computer science. Exploratory factor analysis (EFA), partial least square analysis (PLS) and confirmatory factoranalysis (CFA) was used in this study to generate and select items and establish reliability and validity.

Key words and phrases: Grunderfahrungen des Informatikunterrichts, Scale development, Inventory, Messinstrument, EFA, CFA, PLS.

ZDM Subject Classification: Q20, Q50.

Einleitung

Die Grunderfahrungen des Informatikunterrichts (GI) von Bethge und Fothe (2013) eignen sich einerseits zur Beurteilung der Auswahl von Inhalten und Zielen des Informatikunterrichts aus der Perspektive der Lernenden und andererseits zur Evaluierung des Lernfortschritts, da sie die Erfahrungen und Einsichten während eines Lernprozesses widerspiegeln. Dadurch besteht die Möglichkeit Lernprozesse beim Erwerb von Kompetenzen besser verstehen zu können.

Sie sagen aus, dass der Informatikunterricht dann allgemeinbildend ist, wenn er die drei Grunderfahrungen ermöglicht. Diese sind vielfältig untereinander vernetzt.

- Durch die erste Grunderfahrung (GI1) sollen Informatiksysteme und deren Wirkungen in unterschiedlichen Lebensbereichen entdeckt, verstanden und bewertet werden können. Sie soll die in der Informatik zentrale Rolle der Informationsverarbeitung durch Informatiksysteme widerspiegeln.
- Die zweite Grunderfahrung (GI2) soll das Erkennen ermöglichen, dass Realitätsausschnitte durch Modellierung für ein Informatiksystem aufbereitet werden können, sich Handlungen als Algorithmen formulieren und sich in Programme überführen lassen. Zusätzlich fordert die Grunderfahrung Zwei die Erkenntnis, dass Informatiksysteme von Menschen gestaltet sind.
- Grunderfahrung Drei (GI3) legt den Schwerpunkt auf das Entwickeln von Kompetenzen, welche in der Auseinandersetzung mit Aufgaben und Problemlöse-fähigkeiten innerhalb des Informatikunterrichts und außerhalb der Schule erworben werden sollen.

Da die GI die persönlichen Sichtweisen der Lernenden auf Informatiksysteme und ihre Einstellung und Erfahrungen dazu beschreiben, eignen sie sich für eine psychometrische Messung mit einem Inventar.

Ziele und Hypothesen

Ziel dieser Arbeit ist die Entwicklung eines Fragebogens, um die Grunderfahrungen des Informatikunterrichts von Schülerinnen und Schülern messen zu können und dabei die passende (a) Faktorenstruktur, (b) Reliabilität und (c) Validität zu gewährleisten. (Moosbrugger & Kelava, 2012, S. 8ff) Das Inventar soll einerseits als Messmodell in Strukturgleichungsmodellen und andererseits als Evaluierungsinstrument hinsichtlich der Grunderfahrungen des Informatikunterrichts eingesetzt werden können.

- a) Es ist nicht von vorne herein klar, ob Schülerinnen und Schüler den Informatikunterricht so empfinden wie die Grunderfahrungen des Informatikunterrichts unterteilt sind und sie den Fragebogen im Hinblick auf die Herausbildung einer Drei-Faktoren-Struktur beantworten.
- b) Erwartet wird, dass aus einem Fragenpool in sich konsistente Items gefunden werden können.

- c) Es wurden Korrelationen zu Subskalen von Inventaren, welche die Motivation sowie die Wahrnehmung der eigenen Kompetenz entsprechend der Selbstbestimmungstheorie messen, erwartet (Kiener, 2016). Wegen dieser Korrelationen können Auswirkungen bezüglich der Erfordernisse für eine diskriminante Validität auftreten.

Es bestand die Hypothese, die drei Grunderfahrungen des Informatikunterrichts von Schülerinnen und Schülern mit Hilfe einer psychometrischen Messung unter Einhaltung von Gütekriterien mit ausreichender Reliabilität (innere Konsistenz) und Validität erfassen zu können.

Studienablauf

Um das Ziel ein Messinstrument der GI zu erreichen, wurde zuerst ein Fragenpool von 30 Items, davon 6 negativ formuliert, erstellt und mit zwei iterativ durchgeführten Pretests an 38 Schülerinnen und Schülern eines Gymnasiums mit einem durchschnittlichen Alter von 16 Jahren getestet. Nach den Pretests wurden einige Fragen für die weitere Studie nicht mehr berücksichtigt oder durch Kommentare ergänzt. Im weiteren Verlauf wurde der restliche Fragenpool von 19 Items (2 negativ formuliert) mit Hilfe einer explorativen Faktorenanalyse weiter reduziert (Studie 1). Bei einem Teil der Befragten wurde, um die diskriminante Validität zu untersuchen, ein Unterricht zu Physical-Computing durchgeführt und anschließend zusätzlich zum Fragenpool der GI, ein Fragebogen zur intrinsischen Motivation vorgelegt (Studie 2). Für eine konfirmatorische Faktorenanalyse wurden weitere Daten von Schülerinnen und Schülern erhoben (Studie 3). Die Fragebögen wurde teilweise auf Papier und teilweise als Online-Formular ausgegeben, die Teilnahme erfolgte freiwillig. Um eine möglichst heterogene Stichprobe zu erhalten, wurden auch Probandinnen und Probanden ohne einer Teilnahme an einem Informatikunterricht befragt (Clark & Watson, 1995, S. 313).

Studie 1

Einleitung

Ziel der ersten Teilstudie ist, die Inhaltsvalidität und die Faktorenstruktur des Fragebogens zu den Grunderfahrungen des Informatikunterrichts zu gewährleisten und eine erste Abschätzung der Reliabilität zu erhalten. Dazu wurde das Design

einer Querschnittstudie gewählt. Im Rahmen einer explorativen Faktorenanalyse wurden die Items bewertet und entsprechend der erwarteten Faktorenstruktur ausgewählt.

Methode

Fragenpool zu den Grunderfahrungen des Informatikunterrichts

Zur Sicherstellung der Inhaltsvalidität ist ein Fragenpool anhand der Grunderfahrungen des Informatikunterrichts zu erarbeiten. Dazu ist es notwendig, diese genauer zu betrachten, Hintergründe zu analysieren, sowie Zusammenhänge zu anderen fachdidaktischen Theorien zu finden (Moosbrugger & Kelava, 2012, S. 11-18). Die GI orientieren sich an den Grunderfahrungen des Mathematikunterrichts von Winter, welche sich ebenfalls in drei Abschnitte gliedern. In der ersten Grunderfahrung des Mathematikunterrichts sollen Schülerinnen und Schüler die Erscheinungen der Welt wahrnehmen. In einem Nachsatz wird erläutert, wie Erfahrung zu verstehen ist, nämlich als persönliches Erleben oder sogar als Erleiden von Mathematik (Winter, 1995, S. 37). Obwohl ähnliche Formulierungen in den Grunderfahrungen des Informatikunterrichts fehlen, geben Bethge und Fothe (2013, S. 116) in einer Zusatzbemerkung an, dass es eigentlich um das Staunen geht.

Die Analyse der ersten Grunderfahrung des Informatikunterrichts und des Mathematikunterrichts legen die Formulierung der Items fest. Mit der Wortwahl „Ich“ soll die eigene Wahrnehmung angesprochen werden.

Die Aussagen von GI2 über die Modellierung von Realitätsausschnitten und zur Algorithmisierung von Handlungen wurden sinngemäß für die Formulierung der Items verwendet. Bei Pilottests haben sich Items mit den Wörtern „Algorithmus“ und „Algorithmen“ dahingehend als problematisch erwiesen, dass diese Fragen häufig nicht beantwortet wurden. Eine in Klammern ergänzende Bemerkung „Rechenvorschrift“ hat bessere Ergebnisse erzielt als „Handlungsvorschrift“ oder „Handlungsanleitungen“, obwohl diese Übersetzung nicht in vollem Umfang der Definition von Algorithmus entspricht.

Items zur GI3 sprechen die Einstellung der Probanden zur Anwendbarkeit der erlernten Fähigkeiten und Kompetenzen innerhalb und außerhalb der Informatik an.

Bethge und Fothe (2013, 117) ordneten den Grunderfahrungen Kompetenzen zu, weshalb auch daraus Items formuliert werden können. Das Item „Fehlermeldungen helfen mir im Umgang mit Informatiksystemen“ ist aus der von Magenheimer et al. (2015, S. 41) empirisch ermittelten Kompetenz „Errors as Learning

Opportunities“ abgeleitet worden. Um die Inhaltsvalidität abzusichern, wurde der Fragenpool zur Kontrolle an den Autor der Beiträge zu den Grunderfahrungen des Informatikunterrichts übermittelt.

Datenerhebung und Stichprobe

Der Fragebogen wurde so gestaltet, dass er mit einer fünfstufigen Likert-Skala beantwortet werden kann. Es wurden die Endpunkte der Skala mit „stimme überhaupt nicht zu“ und „stimme voll und ganz zu“ beschriftet, damit eine Intervallskalierung der Daten, eine der Voraussetzungen für die Durchführung einer Faktorenanalyse, gegeben ist (Weijters, Cabooter & Schillewaert, 2010, S. 269) (Bühner, 2011, S. 431-432). Um eine möglichst heterogene Stichprobe zu erhalten wurden die Fragen den Schülerinnen und Schülern mit und ohne besuchten Informatikunterricht vorgelegt. Zusätzlich wurden noch Geschlecht, Alter und Schulstufe sowie die Teilnahme an einem Informatikunterricht erhoben.

Aus pragmatischen Gründen wurde eine Gelegenheitsstichprobe gezogen, sie umfasste 338 (36% männlich, 46% Teilnahme an einem Informatikunterricht) Schülerinnen und Schüler einer neuen Mittelschule, zwei Gymnasien und Studentinnen und Studenten einer pädagogischen Hochschule. Das mittlere Alter betrug 16.94 Jahre ($SD = 3.24$), die Altersspannweite lag zwischen 12 und 25 Jahren.

Datenaufbereitung und Auswertung

Von den 338 rückgemeldeten Fragebögen blieben nach dem Ausschluss von auffälligen, bei denen Antwortmuster erkennbar waren und unvollständig ausgefüllten Fragebögen, 320 übrig. In weiterer Folge wurden davon die 15 extremsten Ausreißer, ermittelt aus der Mahalanobisdistanz, entfernt (Warren, Smith & Cybenko, o.J., S. 14). Die explorative Faktorenanalyse wurde somit mit 305 Datensätzen durchgeführt, womit eine der Voraussetzungen zur Durchführung einer EFA ($n > 300$), erfüllt ist (MacCallum, Widaman, Zhang & Hong, 1999, S. 96). Die explorative Faktorenanalyse sowie die Berechnung der Mahalanobisdistanz, der Bartlett-Test auf Sphärizität, das Kaiser-Mayer-Olkin Kriterium, die Parallelanalyse und die innere Konsistenz wurde in R mit dem Paket „psych 1.8.4“ durchgeführt. Alle weiteren Berechnungen wie Mittelwerte, Standardabweichungen und Korrelationmatritzen erfolgten mit Funktionen, welche in R zur Verfügung stehen.

Variable	M(SD)	α	Schiefe	Exzess	Korr.	KMO
X01	4.55(0.68)	0.83	-1.26	0.51	0.45	0.88
X02	4.38(0.82)	0.83	-1.15	0.45	0.39	0.77
X03	2.89(1.08)	0.82	0.01	-0.67	0.56	0.82
X04	2.53(1.09)	0.82	0.41	-0.48	0.60	0.78
X05	2.51(1.11)	0.82	0.45	-0.48	0.61	0.82
X06	3.22(1.11)	0.82	-0.16	-0.62	0.47	0.88
X07	4.38(0.77)	0.82	-1.16	1.16	0.50	0.84
X08	4.64(0.62)	0.82	-1.73	2.77	0.62	0.86
X09	4.22(0.78)	0.82	-0.62	-0.49	0.56	0.88
X10	4.49(0.70)	0.83	-1.12	0.31	0.34	0.84
X11	3.80(0.98)	0.82	-0.27	-0.66	0.58	0.84
X12	3.76(1.00)	0.82	-0.46	-0.18	0.57	0.82
X13	3.54(1.04)	0.83	-0.46	-0.26	0.45	0.81
X14	3.98(0.91)	0.82	-0.55	-0.33	0.56	0.90
X15	3.13(1.00)	0.84	-0.03	-0.35	0.18	0.53
X16	3.36(1.01)	0.82	-0.26	-0.39	0.50	0.76
X17	3.76(1.02)	0.82	-0.43	-0.54	0.53	0.79
X18	2.99(1.16)	0.82	-0.11	-0.65	0.55	0.84
X19	3.15(1.12)	0.82	-0.10	-0.60	0.53	0.86
Kriterium	<4.5	≤ 0.82	< ± 1.0	< ± 1.0	>0.45	>0.7

Anmerkungen: Werte, welche die Kriterien nicht erfüllen sind kursiv dargestellt. M(SD) = Mittelwert(Standardabweichung). α = Änderung von Cronbachs α , falls die Variable entfernt wird. Korr. = Korrigierte Item-to-total Korrelation. KMO = Kaiser-Mayer-Olkin Kriterium. Kriterium = Auswahlkriterium für die Anforderungen an die Werte.

Tabelle 1. Statistische Kennwerte (n=305)

Ergebnisse

Itemanalyse

Eine weitere Voraussetzung zur Durchführung einer EFA ist, dass die Korrelationen zwischen den Items signifikant von Null abweichen. Der Bartlett-Test auf Sphärizität ist signifikant ($\chi^2(171) = 1747.3, p < .01$) Es kann die Nullhypothese,

dass die Korrelationen zwischen den Items null sind, abgelehnt und somit eine EFA durchgeführt werden.

Eine wesentliche Anforderung an eine EFA ist die Normalverteilung der Items. Das Kriterium dafür ist, dass der Absolutwert der Abweichung von der Schiefe und der Kurtosis kleiner 1 sein soll (Miles & Shevlin, 2001, S. 74). Die Kurtosis einer Normalverteilung ist 3. Die Abweichung von der Kurtosis, der Exzess, ist aus Tabelle 1 zu entnehmen.

Zur Quantifizierung der konvergenten Validität, als Teil der Strukturvalidität, welche ausdrückt, in welchem Maß die unterschiedlichen Messungen des gleichen Konstrukts miteinander korrelieren, wurde die korrigierte Item-to-total Korrelation herangezogen (Park & Kim, 2003, 22). Items sollten mindestens eine Item-to-Total Korrelation 0.4 beziehungsweise 0.5 haben (Gliem & Gliem, 2003) (Bearden & Netemeyer, 1989). Ausgehend davon, dass eine Item-to-Total Korrelation von 0.45 einem 20% Anteil an der gesamten Varianz entspricht, wurde dieser Wert als Kriterium gewählt.

Ein Item (X15) mit einem Kaiser-Mayer-Olkin Kriterium (KMO) unter 0.7 wurde ausgeschlossen, allen anderen Items liegen zwischen 0.78 und 0.91 mit den Bewertungen zur Eignung für eine EFA von „mittel“ (X04, X16) bis „ausgezeichnet“ (X14) (Kaiser, 1974, S. 35). Die Items X01, X02, X10, X13, X15 erhöhen die interne Konsistenz auf 0.83 bzw. 0.84, gemessen durch Cronbachs α , falls sie bei der EFA nicht berücksichtigt werden. Items mit zu hohen Mittelwerten (> 4.5) wurden ebenfalls aus der weiteren Analyse ausgeschlossen, da einerseits die Schiefe der Verteilung zu groß wird und andererseits eine geringe Varianz zu erwarten ist. Mittelwerte, Standardabweichungen, korrigierte Item-to-total Korrelationen, das Kaiser-Mayer-Olkin Kriterium, sowie die Änderung von Cronbachs α , wenn das Item entfernt wird, sind aus Tabelle 1 zu entnehmen. Die für die weitere Analyse ausgewählten Items sind fett gedruckt dargestellt.

Faktorenstruktur

Bei einer durchgeführten Parallelanalyse, bei der die extrahierten Eigenwerte mit Eigenwerten von zufällig generierten Werten verglichen werden, wurden drei Faktoren und drei Hauptkomponenten angezeigt (Horn, 1965, S. 179). Dieses Ergebnis bestätigt den, entsprechend den Grunderfahrungen des Informatikunterrichts, entwickelten Fragenpool.

Die Faktorenanalyse wurde durch die Minimierung der Residuen (minres) und die Rotation der Eigenwerte mit der orthogonalen Rotation (varimax) durchgeführt (Harman & Jones, 1966). Weitere Kriterien zum Ausschluss von Variablen, wie Ladungen auf einen Faktor unter 0.30, und Kommunalitäten unter 0.2, mussten nicht angewendet werden (Harrington, 2009, S. 23) (Child, 2006, S. 47). Aus Tabelle 2 ist zu entnehmen, dass die Variablen X03, X04, X05 auf die erste Grunderfahrung der Informatik (GI1), die Variablen X09, X11, X12, X14 auf die zweite Grunderfahrung der Informatik (GI2) und die Variablen X16, X17, X18, X19 auf die dritte Grunderfahrung der Informatik (GI3) laden. Die drei Faktoren erklären 45% der gesamten Varianz. Die Reliabilitäten, bestimmt über Cronbachs α , betragen für die gesamte Skala ($\alpha = 0.82$), für GI1 ($\alpha = 0.82$), GI2 ($\alpha = 0.70$) und GI3 ($\alpha = 0.68$).

Variable	GI1	GI2	GI3	h^2
X03	0.72	0.15	0.09	0.55
X04	0.82	0.16	0.13	0.71
X05	0.81	0.11	0.20	0.71
X06	0.49	0.10	0.18	0.28
X09	0.14	0.42	0.26	0.26
X11	0.08	0.81	0.05	0.67
X12	0.11	0.66	0.15	0.46
X14	0.16	0.42	0.21	0.25
X16	0.14	0.07	0.63	0.43
X17	0.05	0.21	0.68	0.50
X18	0.29	0.21	0.47	0.35
X19	0.30	0.19	0.40	0.28
Eigenwerte	2.35	1.64	1.47	
erklärte Varianz	20%	13%	12%	45%
Cronbachs α	0.82	0.70	0.68	

Anmerkungen: Ladungen von mehr als 0.4 sind fett gedruckt. h^2 = Kommunalität.

Tabelle 2. Ladungsmatrix und Faktorstruktur (n=305)

Diskussion

Obwohl die Ladungen genügend groß sind ($\lambda \geq 0.4$), weisen die Variablen X06, X09, X14, X18 und X19 lediglich Kommunalitäten $h^2 \leq 0.4$ auf, wodurch eine Stichprobe $n \geq 300$ gerechtfertigt ist (MacCallum et al., 1999). Die Reliabilität liegt mit 0.68 für GI3 unter dem geforderten Grenzwert von 0.7 (Nunnally, 1978, S. 245). Allerdings gilt Cronbachs α als Maß für die Reliabilität nur für τ -äquivalente Messungen, für τ -kongenerische Messungen bei denen die Variablen unterschiedlich auf die Faktoren laden, stellt es eine Abschätzung nach unten dar (Yang & Green, 2011, S. 378).

Studie 2

Einleitung

Ziel der Studie 2 ist, die diskriminante Validität und die konvergente Validität zu prüfen, um sicher zu stellen, dass die Grunderfahrungen des Informatikunterrichts und nicht andere Konstrukte gemessen werden (Moosbrugger & Kelava, 2012, S. 19). In einer vorausgegangenen qualitativen Studie wurden aus Interviews, nach einem Unterricht mit Physical Computing, Kategorien gefunden, welche der Selbstbestimmungstheorie von Deci und Ryan (2008, S. 182) zugeordnet werden konnten (Kiener, 2016, S. 57). Zur Selbstbestimmungstheorie gibt es von Deci und Ryan ein Inventar (IMI), welches als Kurzform (KIM) zur Verfügung steht (Wilde, Bätz, Kovaleva & Urhahne, 2009, S. 31). Die Einschätzung bzw. die Wahrnehmung der eigenen Kompetenz, eine Subskala in beiden Inventaren, könnte mit der Ersten Grunderfahrung des Informatikunterrichts GI1 korrelieren.

Methode

Datenerhebung und Stichprobe

Die Stichprobe ist eine Teilmenge aus Studie 1 mit insgesamt 66 Schülerinnen und Schler aus zwei Gymnasien ($M = 15.6$ Jahre, $SD = 0.96$ Jahre, 46% Jungen). Es wurde eine Interventionsstudie mit einem mehrstündigen Unterricht in Physical Computing durchgeführt (Kiener, 2016, S. 57). Abschließend erfolgte eine Messung der Grunderfahrungen des Informatikunterrichts aus Studie 1 und der Kurzskala intrinsischer Motivation (KIM).

Erhebungsinstrumente

Inventar zu den Grunderfahrungen des Informatikunterrichts mit drei Subskalen und 12 Items aus Studie 1 mit folgenden Reliabilitäten: GI1 ($\alpha = 0.82$), GI2 ($\alpha = 0.70$), GI3 ($\alpha = 0.68$). Die Kurzsкала intrinsischer Motivation (KIM) bestehend aus den Subskalen: Interesse/Vergnügen ($\alpha = 0.85$), Wahrgenommene Kompetenz ($\alpha = 0.83$), Wahrgenommene Wahlfreiheit ($\alpha = 0.75$) und Druck/Anspannung ($\alpha = 0.54$), mit jeweils vier Items pro Subskala. Die Kennwerte weisen auf einen guten Modellfit hin (Wilde et al., 2009, 38).

Datenaufbereitung und Auswertung

Es mussten keine Datensätze aufgrund multivariater Ausreißer entfernt werden. Durch listenweisen Fallausschluss nicht beantworteter Items hat sich die Anzahl der Datensätze auf 64 verringert.

Die Anzahl der Probandinnen und Probanden dieser Studie erfüllt nicht die Voraussetzung zur Durchführung von kovarianz-basierten Verfahren. Allerdings sind mit Partial Least Square (PLS), Analysen bei mittleren Faktorladungen und Stichproben ab $n \geq 50$ möglich (Chin & Newsted, 2000, 333). PLS ist ein komponentenbasiertes Verfahren und die Parallelanalyse aus Studie 1 hat gezeigt, dass die Anzahl der Faktoren denen der Hauptkomponenten entspricht (Chin, 2014, 301).

In PLS steht mit dem „Heterotrait-Monotrait Ratio“ (HTMT) ein Maß zur Beurteilung der diskriminanten Validität zur Verfügung. Liegt der HTMT-Wert nahe bei Eins, muss davon ausgegangen werden, dass die Skalen das gleiche Konstrukt messen. Als Schwelle dafür werden HTMT-Werte von 0.85 bis 0.90 genannt (Henseler, Ringle & Sarstedt, 2015).

Zur Analyse der konvergenten Validität wird die „Average Variance Extracted“ (AVE) herangezogen. Für eine ausreichende Validität werden AVE-Werte von mindestens 0.5 gefordert. Geringfügige Unterschreitungen sind möglich, falls die „Composit Reliability“ (CR) genügend groß ist. Die Quadratwurzel aus AVE sollte größer als die Korrelationen zwischen den Konstrukten sein (Fornell & Larcker, 1981) (Chin, 2010, S. 671).

Die Auswertung wurde in R mit dem Paket „matrixpls 1.0.5“ durchgeführt. Mittelwerte, Standardabweichungen und Kovarianzmatrizen wurden mit Funktionen von R berechnet.

Ergebnisse

Die konfirmatorisch geführte Analyse mit PLS liefert folgende Ladungen der Prädiktoren auf GI1: X03 ($\lambda = 0.87$), X04 ($\lambda = 0.90$), X05 ($\lambda = 0.88$), X06 ($\lambda = 0.72$), auf GI2: X09 ($\lambda = 0.70$), X11 ($\lambda = 0.75$), X12 ($\lambda = 0.71$), X14 ($\lambda = 0.69$) und auf GI3: X16 ($\lambda = 0.72$), X17 ($\lambda = 0.72$), X18 ($\lambda = 0.68$), X19 ($\lambda = 0.65$). Die HTMT-Werte der Analyse sind in Tabelle 3 angeführt. Die Interkonstrukt-Korrelationen stehen in Klammern.

Variable	GI1	GI2	GI3
GI1	0		
GI2	0.26 (0.19)	0	
GI3	0.52 (0.37)*	0.29 (0.20)	0
Interesse	0.57 (0.47)**	-0.05 (-0.05)	0.26 (0.19)
Kompetenz	0.61 (0.43)**	0.15 (0.10)	0.34 (0.23)
Wahlfreiheit	0.48 (0.29)*	0.32 (0.18)	0.46 (0.25)*
Druck	-0.28 (0.11)	-0.42 (-0.29)*	0.13 (0.09)
AVE/ \sqrt{AVE}	0.61/0.78	0.51/0.71	0.48/0.6
CR	0.85	0.82	0.79

Anmerkungen: () Pearson-Korrelationen zwischen den Faktoren. AVE = Average Variance Extracted. CR = Composite Reliability. * $p < 0.05$, ** $p < 0.001$.

Tabelle 3. HTMT- und Korrelations-Matrix (n=64)

Diskussion

Wie erwartet liefert PLS höhere Ladungen als bei einer kovarianz-basierten Faktorenanalyse, da bei der Extraktion mit Hilfe der Hauptkomponenten die Residuen in die Varianz eingerechnet werden (Widaman, 1993, S. 285).

Aus Tabelle 3 ist zu entnehmen, dass kein HTMT-Wert nahe bei 0.85 oder darüber liegt, womit davon ausgegangen werden kann, dass die latenten Variablen jeweils unterschiedliche Konstrukte messen und daher eine ausreichende diskriminante Validität angenommen werden kann. Zu bemerken ist, dass aus der KIM die Subskalen „Wahrnehmung der eigenen Kompetenz“ mit 0.43 ($p < 0.001$) und

„Interesse“ mit 0.47 ($p < 0.001$) die höchsten Korrelationen zu GI1 aufweisen, da beide Instrumente Aspekte der Kompetenz und des Interesses messen. Die Average Variance Extracted (AVE) des Konstrukts GI3 unterschreitet mit 0.48 geringfügig die Schwelle von 0.5, welches aber bei einer Composite Reliability (CR) von 0.79 vertretbar ist, siehe Tabelle 3 (Chin, 2010). Die Quadratwurzel von AVE der Konstrukte ist in allen Fällen größer als die Interkonstrukt-Korrelation.

Studie 3

Einleitung

Um die Prüfung der Reliabilität und Validität der Skalen zur Messung der Grunderfahrungen des Informatikunterrichts abzuschließen, wurde eine neue Stichprobe erhoben. Da die in Studie 1 gefundene Faktorenstruktur ausschließlich auf Kennwerten beruhte, war es notwendig, die faktorielle Validität als Teil der Konstruktvalidität mit der konfirmatorischen Faktorenanalyse (CFA) zu untersuchen. Wesentlich dabei ist, dass zusätzlich zu den Fitparametern noch eine inferenzstatistische Prüfung der Modelle zur Anwendung kommt.

Methode

Datenerhebung und Stichprobe

Für die Studie 3 wurde ein halbes Jahr nach der Datenerhebung von Studie 1 an zwei Gymnasien eine Gelegenheitsstichprobe von 218 (48% Mädchen) Schülerinnen und Schülern gezogen. Verwendet wurde der Fragebogen mit den in Studie 1 ermittelten Items. Zusätzlich wurde das Alter ($M = 14,25$ Jahre, $SD = 1,28$ Jahre) und die Teilnahme an einem Informatikunterricht (50,5%) erhoben. Der Fragebogen wurde von den Teilnehmerinnen und Teilnehmern zur Gänze online ausgefüllt. Die Teilnahme wurde vom jeweiligen Klassenlehrer organisiert, wobei die Freiwilligkeit gegeben war. Die Stichprobengröße wurde mit $n = 200$ festgelegt. Die Richtwerte einer minimalen Stichprobengröße zur Durchführung einer CFA liegen zwischen 150 bei Normalverteilung und 200 bei einer Verletzung der Normalverteilung und einer mittleren Ladung von 0.4 (Muthén & Muthén, 2002) (Jackson, Voth & Frey, 2013).

Datenaufbereitung und Auswertung

Alle Datensätze konnten für die Studie verwendet werden, da die Fragen zwingend beantwortet werden mussten. Es wurden 18 multivariate Ausreißer, bestimmt aus der Mahalanobisdistanz, aus der Stichprobe entfernt. Basierend auf der in Studie 1 gefundenen Faktorenstruktur, mit den drei latenten Konstrukten GI1, GI2 und GI3 und den insgesamt zwölf Prädiktoren, wurde in R mit dem Paket „Lavaan 0.6-1“ eine CFA durchgeführt. Als Schätzer wurde Robust Maximum Likelihood (RML) gewählt, da dieses Verfahren für ordinale Daten und bei mäßigen Verletzungen der Normalverteilung geeignet ist (Li, 2016) (Viladrich, Angulo-Brunet & Doval, 2017).

Der Mardia-Test auf multivariate Schiefe ist signifikant, obwohl die univariaten Werte der Verteilungen im akzeptablen Bereich liegen ($-0.95 < \text{Schiefe} < 0.24$, $-0.7 < \text{Excess} < 0.26$) außer bei X11, hier beträgt der Excess = -1.22. Deshalb wurden die p-Werte für die Chiquadrate der Modelle über Bootstrap nach Bollen und Stine ermittelt (Bühner, 2011, 432). Es wurden zwei genestete Modelle, welche sich durch die Residualvarianzen der Prädiktoren unterscheiden, miteinander verglichen. In Modell A sind alle Residualvarianzen auf Null fixiert, in Modell B ist eine Residualvarianz zwischen X11 und X12 modelliert. Diese Residualvarianz ist inhaltlich durch das Wort „Problemlösefähigkeiten“ in X11 und X12 begründet. Die Modifikationsindizes weisen auf eine weitere Verbesserung der Modellanpassung durch eine Residualvarianz zwischen X11 und X12 an, begründet durch das Vorkommen von „Algorithmen“ in beiden Items. Diese wurde aber nicht modelliert, da diese Items die Meinung der Schülerinnen und Schüler zu Algorithmen messen sollen. Zusätzlich wurde ein Modell (C) mit neun Prädiktoren untersucht, darin wurden jene Items mit den geringsten Kommunalitäten aus Studie 1 entfernt: X06 ($h^2=0.28$), X14 ($h^2=0.25$) und X19 ($h^2=0.28$). Alle Modellierungen erfolgten kongenerisch und die Varianzen der Faktoren GI1, GI2 und GI3 wurden auf eins fixiert.

Ergebnisse

Die Chiquadrate der Modelle, ihre p-Werte sowie die Anpassungsindizes sind in Tabelle 4 dargestellt. Die Modelle sind absteigend nach den AIC und BIC Werten sortiert. Model A ($\chi^2(51) = 90.5$, $p = .009$; $\chi^2/\text{df}=1.77$; SRMR=.076; RMSEA = .062, $p=.159$, 90% CI [.041, .083]; CFI=.941) zeigt beim p-Wert von χ^2 eine signifikant abweichende Modellanpassung an und auch der CFI erreicht nicht die erforderlichen Werte für einen mindestens akzeptablen Modellfit. Modell

B ($\chi^2(50) = 73.1$, $p = .015$; $\chi^2/df=1.46$; SRMR=.061; RMSEA = .048, $p=.531$, 90% CI [.041, .07]; CFI=.965) erreicht einen akzeptablen Modellfit. Modell C ($\chi^2(24) = 36.3$, $p = .058$; $\chi^2/df=1.5$; SRMR=.07; RMSEA = .051, $p=.453$, 90% CI [.00, .083]; CFI=.977) erreicht für χ^2 , RMSEA und CFI gute Werte und für SRMR einen akzeptablen Wert der Modellanpassung.

Modell	$\chi^2(p)$	df	χ^2/df	SRMR	RMSEA	CI	CFI	AIC	BIC
A	90.5(.009)	51	1.77	.076	.062**	.04	.941	6078	6167
B	73.1(.015)	50	1.46	.061	.048**	.02	.965	6062	6154
C	36.3(.058)	24	1.50	.070	.051**	.00	.977	4504	4573
akzept.	($\geq .01$)		≤ 3.0	$\leq .10$	$\leq .08^*$		$\geq .95$		
gut	($\geq .05$)		≤ 2.0	$\leq .05$	$\leq .05^{**}$.00	$\geq .97$		

Anmerkungen: χ^2 = Chiquadrat des Modells. p = p -Wert von χ^2 (Bollen-Stine Bootstrap). df = Freiheitsgrade. RMSEA = Root Mean Square Error of Approximation. p -Werte für RMSEA $\leq .05$ (* $p \geq 0.05$, ** $p \geq 0.10$). CI = linker Rand des 90% Konfidenzintervalls von RMSEA. CFI = Bentler Comparative Fit Index. SRMR = Standardized Root Mean Square Residual. AIC = Akaike Information Criterion. BIC = Bayesian Information Criterion.

Table 4. Modellgüte und Gütekriterien der getesteten Modelle ($n=218$)

Die Ladungsmatrix mit den geschätzten Faktorladungen (Est.), deren standardisierter Fehler (S.E.) sowie die standardisierten Ladungsgewichte (Std.) und die Reliabilitäten, ausgedrückt durch Raykovs ρ , sind für die Modelle B und C in Tabelle 5 dargestellt. Alle Items laden signifikant ($p < 0.001$) auf die Faktoren. Die Kovarianz, deren standardisierter Fehler und die Residualkorrelation zwischen den Items X16 und X17 sind nicht in der Tabelle angeführt. (Kovarianz = 0.32, S.E.= 0.07, Residualkorrelation = 0.46) Im unteren Teil der Tabelle 5 sind die standardisierten Korrelationen (Est. und Std.) und deren standardisierte Fehler (S.E.) zwischen den latenten Konstrukten GI1, GI2 und GI3 abzulesen. In Modell B lädt X16 mit dem geringsten Wert auf einen Faktor (GI3, $\lambda=.44$, $R^2=.194$) und trägt immerhin noch mit einem gerundeten Varianzanteil von 20% an der Varianz bei. Die Reliabilitäten der Subskalen betragen .84 (GI1), .72 (GI2) und .71 (GI3). Obwohl für Modell B ein akzeptabler Modellfit erreicht wurde, ist eine signifikante lokale Abweichung der Korrelationsmatrix zwischen den Items X09 und X19 mit der standardisierten Abweichung von 3.46 ($p < .01$) zu erkennen. Modell B unterschätzt die Korrelation zwischen X09 und X19 signifikant. Modell C zeigt im Mittel höhere Ladungen der Items als Modell B, mit dem geringsten Wert für

X09 ($\lambda=.51$, $R^2=.26$). Die Reliabilitäten der Subskalen haben sich zu Modell B nicht wesentlich geändert, Raykovs ρ beträgt .87 (GI1), .70 (GI2) und .73 (GI3). In Modell C weisen die Residuen darauf hin, dass die Korrelationen zwischen X09 und X17 mit 3.03 ($p < .01$) und X09 und X18 mit 3.659 ($p < .01$) signifikant unterschätzt und zwischen X17 und X18 mit -4.664 ($p < .01$) signifikant überschätzt werden.

Item	Skala	Modell B				Modell C			
		Est.	S.E.	Std.	ρ	Est.	S.E.	Std.	ρ
X03	GI1	.89	.07	.81	.84	.87	.07	.79	.87
X04	GI1	.95	.06	.89		.96	.06	.90	
X05	GI1	.89	.06	.81		.89	.07	.81	
X06	GI1	.52	.09	.48					
X09	GI2	.49	.08	.60	.72	.41	.07	.51	.70
X11	GI2	.60	.07	.69		.69	.06	.79	
X12	GI2	.56	.08	.65		.57	.06	.66	
X14	GI2	.46	.06	.55					
X16	GI3	.49	.11	.44	.71	.75	.09	.67	.73
X17	GI3	.63	.08	.67		.82	.07	.88	
X18	GI3	.72	.09	.68		.58	.09	.55	
X19	GI3	.62	.09	.60					
GI1	GI2	.43	.09	.43		.40	.06	.40	
GI1	GI3	.42	.08	.42		.28	.08	.28	
GI2	GI3	.39	.14	.39		.22	.09	.22	
Gesamt					.86				.85

Anmerkungen: Est. = geschätzte Ladungsgewichte (Faktorladungen). S.E. = Standardfehler Fehler der Ladungsgewichte. Std. = Standardisierte Ladungsgewichte, ρ = Raykovs ρ

Tabelle 5. Ladungen, Varianzen und Reliabilitäten der Modelle (n=218)

Diskussion

Ziel der Studie 3 war, die in Studie 1 gefundene Faktorenstruktur mit einem unabhängigen Datensatz zu überprüfen. Die Anpassungsindizes der CFA zeigen, dass die empirische Varianz-Kovarianzmatrix von Modell B akzeptabel und von Modell C gut reproduziert werden kann. Die Reliabilitäten beider Modelle erreichen die Mindestanforderungen für eine reliable Skala von 0.7 (Nunnally, 1978). Die Ergebnisse unterstützen die Evidenz einer Drei-Faktorenstruktur der Grunderfahrungen des Informatikunterrichts. Zu erwarten war, dass sich bei einer Reduzierung der Items in Modell C die Reliabilitäten verringern. Sie haben sich aber für die Subskalen GI1 und GI3, durch Wegnahme der Items X06 und X19, erhöht, da sie wenig zur Varianzaufklärung beitragen.

Limitierungen

Mit drei Items pro Faktor in Modell C wurde die unterste Grenze für ein Inventar zur psychometrischen Messung erreicht. Wegen der daraus resultierenden geraden Identifizierung einer Subskala, sollte die Kurzsкала (Modell C) in Strukturgleichungsmodellen nur mit allen drei Faktoren verwendet werden. Die Grunderfahrungen Eins eignen sich gut zur Formulierung von Items für eine Likertskala, erkennbar an den guten psychometrischen Eigenschaften der Skala. Die Aussagen der Grunderfahrungen Zwei des Informatikunterrichts gestalteten sich für eine Zustimmungsskala problematisch, da hier nicht die Einstellungen adressiert werden, sondern die persönlichen Sichtweisen auf Fakten. Ein Bedarf besteht für eine Ergänzung und Weiterentwicklung der Items bezüglich der Subskalen GI2 und GI3. Bei der Pilotierung und bei der Untersuchung der Items in Studie 1 hat sich gezeigt, dass invers formulierte Item nicht geeignet waren, wodurch sich der ursprüngliche Itempool stark reduziert hat. Zur Verbesserung des Inventars sollte das verbliebene invers formulierte Items ersetzt werden.

Literatur

- Bearden, W. O. & Netemeyer. (1989). Measurement of consumer susceptibility to interpersonal influence. *Journal of Consumer Research*, 15 (4), 473–481. doi: 10.1002/mar.4220090504
- Bethge, B. & Fothe, M. (2013). Grunderfahrungen des informatikunterrichts - ein beitrag zur frage der allgemeinbidung von informatik. In N. Breier,

- P. Stechert & T. Wilke (Hrsg.), *Infos 2013* (S. 113–121). Kiel: Dept. of Computer Science Kiel Univ.
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (3., aktualisierte und erw. Aufl.). München: Pearson Studium. doi: 10.17877/DE290R-6148
- Child, D. (2006). *The essentials of factor analysis* (3. Aufl.). London: Continuum.
- Chin, W. W. (2010). How to write up and report pls analyses. In V. Esposito Vinzi (Hrsg.), *Handbook of partial least squares* (S. 655–690). Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg.
- Chin, W. W. (2014). The partial least square approach to structural equation modeling. In G. A. Marcoulides (Hrsg.), *Modern methods for business research* (S. 295–336). London: Psychology Press.
- Chin, W. W. & Newsted, P. R. (2000). Structural equation modeling analysis with small samples using partial least squares. In R. H. Hoyle (Hrsg.), *Statistical strategies for small sample research* (S. 307–341). Thousand Oaks, Calif.: Sage Publ. doi: 10.1007/978-3-540-32827-829
- Clark, L. A. & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, Psychol Assess*, 7 (3), 309–319. doi: 10.1037/1040-3590.7.3.309
- Deci, E. L. & Ryan, R. M. (2008). Self-determination theory: A macrotheory of human motivation, development, and health. *Canadian Psychology/Psychologie canadienne, Can Psychol*, 49 (3), 182–185. doi: 10.1037/a0012801
- Fornell, C. & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18 (1), 39. doi: 10.2307/3151312
- Gliem, J. A. & Gliem, R. R. (2003). Calculating, interpreting, and reporting cronbach's alpha reliability coefficient for likert-type scales. *Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education*, 82–88.
- Harman, H. H. & Jones, W. H. (1966). Factor analysis by minimizing residuals (minres). *Psychometrika*, 31 (3), 351–368. doi: 10.1007/BF02289468
- Harrington, D. (2009). *Confirmatory factor analysis*. Oxford and New York: Oxford University Press.
- Henseler, J., Ringle, C. M. & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the academy of marketing science, J Acad Mark Sci*, 43 (1), 115–135.

doi: 10.1007/s11747-014-0403-8

- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30 (2), 179–185. doi: 10.1007/BF02289447
- Jackson, D. L., Voth, J. & Frey, M. P. (2013). A note on sample size and solution propriety for confirmatory factor analytic models. *Structural equation modeling: a multidisciplinary journal*, 20 (1), 86–97. doi: 10.1080/10705511.2013.742388
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39 (1), 31–36. doi: 10.1007/BF02291575
- Kiener, A. (2016). Physical computing im informatikunterricht-eindrücke und erfahrungen von schülerinnen und schülern. *LOG IN*, 36 (185/186), 54–59.
- Li, C.-H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior research methods*, 48 (3), 936–949. doi: 10.3758/s13428-015-0619-7
- MacCallum, R. C., Widaman, K. F., Zhang, S. & Hong, S. (1999). Sample size in factor analysis. *Psychological methods, Psychol Methods*, 4 (1), 84–99. doi: 10.1037/1082-989X.4.1.84
- Magenheim, J., Schubert, S. & Schaper, N. (2015). Modelling and measurement of competencies in computer science education. *KEYCIT 2014: key competencies in informatics and ICT, KEYCIT 2014*, 7 (1), 33–57.
- Miles, J. & Shevlin, M. (2001). *Applying regression and correlation: A guide for students and researchers*. London: Sage Publ.
- Moosbrugger, H. & Kelava, A. (2012). *Testtheorie und fragebogenkonstruktion* (2., aktualisierte und überarbeitete Auflage Aufl.). Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg. doi: 10.1007/978-3-642-20072-4
- Muthén, L. K. & Muthén, B. O. (2002). How to use a monte carlo study to decide on sample size and determine power. *Structural equation modeling: a multidisciplinary journal*, 9 (4), 599–620. doi: 10.1207/S15328007SEM0904”_-”8
- Nunnally, J. C. (1978). *Psychometric theory* (2. ed. Aufl.). New York: McGraw-Hill.
- Park, C.-H. & Kim, Y.-G. (2003). Identifying key factors affecting consumer purchase behavior in an online shopping context. *International journal of retail and distribution management, IJRDM*, 31 (1), 16–29. doi: 10.1108/09590550310457818
- Viladrich, C., Angulo-Brunet, A. & Doval, E. (2017). Un viaje alrededor de

- alfa y omega para estimar la fiabilidad de consistencia interna. *Anales de Psicología*, 33 (3), 755–782. doi: 10.6018/analesps.33.3.268401
- Warren, R., Smith, R. F. & Cybenko, A. K. (o. J.). *Use of mahalanobis distance for detecting outliers and outlier clusters in markedly non-normal data: a vehicular traffic example.*
- Weijters, B., Cabooter, E. & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing, IJRM*, 27 (3), 236–247. doi: 10.1016/j.ijresmar.2010.02.004
- Widaman, K. F. (1993). Common factor analysis versus principal component analysis: Differential bias in representing model parameters? *Multivariate Behav Res*, 28 (3), 263–311.
- Wilde, M., Bätz, K., Kovaleva, A. & Urhahne, D. (2009). Überprüfung einer kurzskala intrinsischer motivation (kim). *Zeitschrift für Didaktik der Naturwissenschaften (ZfDN)*, 15, 31–45.
- Winter, H. (1995). Mathematikunterricht und allgemeinbildung. *Mitteilungen der Gesellschaft für Didaktik der Mathematik, Mitteilungen der GDM*, 21 (61), 37–46. doi: 10.1515/dmvm-1996-0214
- Yang, Y. & Green, S. B. (2011). Coefficient alpha: A reliability coefficient for the 21st century? *Journal of Psychoeducational Assessment, JPA*, 29 (4), 377–392. doi: 10.1177/0734282911406668

ANDREAS KIENER
PRIVATE PÄDAGOGISCHE HOCHSCHULE DER DIÖZESE LINZ
SALESIANUMWEG 3, 4020 LINZ

PARIS LODRON UNIVERSITÄT SALZBURG
HELLBRUNNER STRASSE 34, 5020 SALZBURG

E-mail: andreas.kiener@ph-linz.at

(Received March, 2019)

Anhang A. Anhang

Nummer	Item
X01	Informatiksysteme sind in vielen technischen Geräten vorhanden.
X02	Selbst kleinste Informatiksysteme können sehr komplex sein.
X03	Ich erkenne den Grundaufbau von Informatiksystemen in Alltagsgeräten.
X04	Ich kann wesentliche Bestandteile von Informatiksystemen benennen.
X05	Ich kann das grundlegende Prinzip von Informatiksystemen erläutern.
X06	Ich habe Probleme die grundlegende Funktionsweise von Informatiksystemen zu verstehen.
X07	Es gibt viele Anwendungsmöglichkeiten für Informatiksysteme.
X08	Informatiksysteme können sehr nützlich sein.
X09	Informatiksysteme können immer mehr Aufgaben von Menschen übernehmen.
X10	Informatiksysteme sind von Menschen erschaffen.
X11	Handlungen können als Algorithmen (Rechenvorschriften) formuliert werden.
X12	Informatiksysteme werden von Algorithmen (Rechenvorschriften) gesteuert.
X13	Die Daten der realen Welt müssen für ein Informatiksystem aufbereitet werden.
X14	Programme enthalten die Befehle zur Steuerung von Informatiksystemen.
X15	Es gibt innerhalb der Informatik nur wenige Anwendungsmöglichkeiten der im Informatikunterricht erworbenen Problemlösefähigkeiten.
X16	Die im Informatikunterricht erworbenen Problemlösefähigkeiten sind auch in anderen Fächern anwendbar.
X17	Die im Informatikunterricht erworbenen Problemlösefähigkeiten können außerhalb der Schule sehr nützlich sein.
X18	Fehlermeldungen helfen mir im Umgang mit Informatiksystemen.
X19	Für fast jedes Problem gibt es zur Lösung ein entsprechendes Informatiksystem.

Zuordnung der Items zu den Subskalen GI1, GI2 und GI3 der Grunderfahrungen des Informatikunterrichts GI1: X03, X04, X05, (X06 R); GI2: X09, X11, X12, (X14); GI3: X16, X17, X18, (X19). Die in Klammern stehenden Items sind optional. Endpunktbewertete fünfstufige Zustimmungsskala von „stimme überhaupt nicht zu“ bis zu „stimme voll und ganz zu“.