



Untersuchung des Inventars zu den Grunderfahrungen des Informatikunterrichts auf Messinvarianz zwischen Lernenden mit und ohne besuchten Informatikunterricht und Schülerinnen und Schülern

Andreas Kiener

Pädagogische Hochschule der Diözese Linz
andreas.kiener@ph-linz.at

EINGEREICHT 31 MÄRZ, 2020

ÜBERARBEITET 10 AUG, 2020

ANGENOMMEN 2 OKT, 2020

Mit den drei Grunderfahrungen des Informatikunterrichts, welche mit der Wahrnehmung der eigenen Kompetenz, der Bedeutung von Algorithmen für Informatiksysteme und der Anwendbarkeit der erlernten Kompetenzen umschrieben werden können, ist es möglich Lernprozesse in der Auseinandersetzung mit Informatiksystemen zu untersuchen. Mit dem Inventar zu den Grunderfahrungen des Informatikunterrichts IGI können sie auch gemessen und quantifiziert werden. Anhand einer Gelegenheitsstichprobe von 428 Schülerinnen und Schülern aus drei Gymnasien und einer Neuen Mittelschule konnte zumindest eine partielle skalare Messinvarianz zwischen Mädchen und Jungen und Lernenden mit und ohne besuchten Informatikunterricht nachgewiesen werden. Dadurch war es möglich die latenten Mittelwerte der drei Skalen des Inventars zwischen den Gruppen zu vergleichen. Im Ergebnis zeigten sich für die erste und zweite Grunderfahrung geringe bis mittlere Effekte und für die dritte Grunderfahrung keine signifikanten Unterschiede. Dafür stellte sich für die dritte Grunderfahrung eine geringe negative Abhängigkeit vom Alter heraus.

SCHLÜSSELWÖRTER: Inventar, Messinvarianz, Grunderfahrungen des Informatikunterrichts

1. Einleitung

Mit dem Inventar zu den Grunderfahrungen des Informatikunterrichts (IGI) ist es möglich, Unterschiede in der persönlichen Entwicklung von Schülerinnen und Schülern im Vergleich zwischen jenen mit einem besuchten Informatikunterricht und jenen ohne Informatikunterricht zu untersuchen und zu quantifizieren (Kiener, 2019).

Die Grunderfahrungen des Informatikunterrichts von Bethge und Fothe (2013, 2014) sagen aus, dass der Informatikunterricht dann allgemeinbildend ist, wenn er die folgenden drei Grunderfahrungen im Unterricht ermöglicht:

- 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 es ermöglichen zu erkennen, dass für ein Informatiksystem Realitätsausschnitte durch Modellierung aufbereitet werden müssen, sich Handlungen als Algorithmen formulieren und in Programme überführen lassen.
- Grunderfahrung Drei (GI3) beleuchtet die Entwicklung der Einstellungen, welche in der Auseinandersetzung mit Aufgaben und Problemstellungen innerhalb des Informatikunterrichts und außerhalb der Schule erworben und angewendet werden können.

Es ist nicht von vornherein gesichert, dass das Inventar bei Schülerinnen und Schülern ohne einen besuchten Informatikunterricht das gleiche Konstrukt misst, wie bei jenen mit einem länger besuchten Informatikunterricht. Diese Tatsache liegt darin begründet, dass die Grunderfahrungen des Informatikunterrichts den persönlichen Entwicklungsstand und die persönliche Einstellung zur Informatik betrachten. Somit kann das Antwortverhalten bezüglich einzelner Items derart verändert sein, dass sich die persönliche Entwicklung nicht nur über eine graduelle Veränderung auf der Zustimmungsskala auswirkt, sondern dass die Aussage eines Items aufgrund neuer Erkenntnisse aus einem anderen Blickwinkel gesehen und deswegen unterschiedlich bewertet wird. Die Folge ist, dass das Item auf eine andere Subskala oder einer nicht modellierten Variablen lädt. Somit wäre das Inventar zu den Grunderfahrungen des Informatikunterrichts zwischen den Gruppen nicht mehr messinvariant bezüglich der Gruppe, welche keinen Informatikunterricht besucht hat und der Gruppe mit einem besuchten Informatikunterricht (van de Schoot, Lugtig & Hox, 2012). Deshalb ist es notwendig das Inventar IGI auf Messinvarianz zu prüfen.

Allerdings ist der Informatikunterricht nicht der einzige Lernort, der den Einfluss eines Informatikunterrichts auf Lernende verzerren kann. Neben dem Informatikunterricht beeinflusst auch die Computernutzung zu Hause die Leistungen positiv (Ponzo, 2011).

In einer Untersuchung von Tablet-Klassen an Gymnasien wurde auch erhoben, dass computerbezogene Kompetenzen überwiegend außerschulisch erlernt werden. Allerdings ist der Anteil der Schülerinnen und Schüler, welche die Kompetenzen in der Schule erwerben, bei Mädchen größer als bei Jungen (Massek, Drossel & Eickelmann, 2018).

Zusätzlich scheint es, dass auch das Alter wegen der daraus implizierten längeren Auseinandersetzung mit Computern und Informatiksystemen die Grunderfahrungen des Informatikunterrichts beeinflussen könnte. Mit dem Inventar INCOBI-R, der revidierten Fassung des INCOBI, ließen sich die computerbezogenen

Einstellungen von Lernenden erfassen. Die Einstellungen betreffen die persönlichen Erfahrungen bezüglich der Nützlichkeit von Computern und der Ansicht, ob Computer eine unbeeinflussbare Technik darstellen. Die Korrelation der computerbezogenen Einstellung mit den Jahren der Computernutzung wurde mit rund $r = 0.24$ bestimmt (Richter, Naumann & Horz, 2010).

Es bestanden zudem Unterschiede in allen computerbezogenen Variablen in dem für die Untersuchung verwendeten INCOBI, wobei sie bei computerbezogenen Einstellungen am geringsten waren. Die geschlechtsspezifischen Unterschiede ließen sich zum großen Teil auf die Dauer der Computernutzung zurückführen (Richter, Naumann & Horz, 2002).

Jungen lagen bei einer Untersuchung über das Selbstbewusstsein im Umgang mit Rechnern mit 2.95 (von max. 4) Punkten signifikant höher als Mädchen, welche 2.4 erreichten (Magenheim & Schulte, 2005). Geschlechtsspezifische Unterschiede in den „Computer science skills“ sind sowohl in China, Indien, Russland und den Vereinigten Staaten zu beobachten (Loyalka et al., 2019).

Dem gegenüber stehen die Ergebnisse der ICILS 2018 Studie zu Computational Thinking, bei der die Mädchen bei Leistungstests einen Leistungsvorsprung in allen untersuchten Ländern aufwiesen (Eickelmann et al., 2019).

Wegen der teils widersprüchlichen Ergebnisse der Studien zu computerbezogenen Einstellungen und Leistungen der informatischen Kompetenzen zwischen Mädchen und Jungen ist es sinnvoll, die Messinvarianz auch zwischen den Geschlechtern zu überprüfen und die Unterschiede mit Hilfe des IGI zu untersuchen und zu quantifizieren.

Wegen der zuvor beschriebenen Unterschiede zwischen Lernenden bezüglich der computerbezogenen Einstellungen und dem Geschlecht ist das Ziel dieser Arbeit, die Messinvarianz des IGI zwischen diesen Gruppen zu überprüfen. Falls keine Messinvarianz oder nur eine partielle Messinvarianz nachgewiesen werden kann, sollen diese Items auch identifiziert werden. In weiterer Folge können dann die Mittelwerte der Skalen des IGI zwischen den Gruppen verglichen werden.

Fragestellungen

- a) Erfassen die Skalen des Inventars der Grunderfahrungen des Informatikunterrichts IGI bei beiden Geschlechtern und bei Lernenden mit und ohne Informatikunterricht das gleiche Konstrukt und zeigt das IGI zumindest skalare Invarianz?
- b) Gibt es Mittelwertsunterschiede in den latenten Konstrukten der Grunderfahrungen des Informatikunterrichts zwischen Schülerinnen und Schülern?
- c) Gibt es Auswirkungen bezüglich des Ausmaßes eines Informatikunterrichts auf die latenten Mittelwerte der Grunderfahrungen des Informatikunterrichts?
- d) Gibt es einen Zusammenhang der Grunderfahrungen des Informatikunterrichts mit dem Alter?

2. Methode

2.1 Forschungsdesign

Um die Unterschiede in der mittleren Ausprägung der latenten Konstrukte der Grunderfahrungen des Informatikunterrichts bewerten zu können, muss das Inventar über die Gruppen hinweg das gleiche Konstrukt messen. Deshalb wird zuerst Punkt (a) der Fragestellungen untersucht.

Zu deren Klärung können mit Hilfe der Messinvarianz die Restriktionen der Modelle in den Ladungen und Achsenabschnitte (Intercepts) zwischen den Gruppen stufenweise erhöht werden. Die Stufen unterscheiden sich dadurch, welche Parameter gruppenübergreifend fixiert werden. Folgende Stufen der Messinvarianz wurden untersucht:

1. Konfigurale Invarianz ($M_1 = \text{Modell Gender}$, $MI_1 = \text{Modell Informatikunterricht}$): Die manifesten Variablen (Items) laden alle auf die gleichen latenten Variablen, den Subskalen des IGI. Das Modell hat in beiden Gruppen die gleiche Struktur bzw. Konfiguration.
2. Metrische Invarianz (M_2 , MI_2): Voraussetzung ist die konfigurale Messinvarianz. Die Ladungen der manifesten Variablen werden in beiden Gruppen fixiert, das bedeutet, dass somit zusätzlich die Konstruktvalidität gefordert wird.
3. Skalare Invarianz (M_3 , MI_3): Voraussetzung ist die metrische Invarianz. Die Intercepts der manifesten Variablen werden gruppenübergreifend gleichgesetzt. Dadurch können Mittelwertsunterschiede in den latenten Variablen untersucht werden, da deren Mittelwerte auf die Mittelwerte der manifesten Variablen zurückzuführen sind (Vandenberg & Lance, 2000).

Falls nur ein partial invariantes Modell (M_4 , MI_4) gefunden werden kann, so ist auch die Identifizierung von nicht invarianten manifesten Variablen von Interesse. Daraus wird ein Einblick in die unterschiedlichen Auffassungen der Gruppen bezüglich dieser Items erlangt.

Um Mittelwerte latenter Variablen vergleichen zu können, ist zumindest eine skalare oder partial skalare Invarianz notwendig (Milfont & Fischer, 2010; Putnick & Bornstein, 2016). Für eine partial skalare Invarianz müssen mindestens zwei manifeste Variablen mit ihren Ladungen und Intercepts invariant sein (Steenkamp & Baumgartner, 1998). Für das zu untersuchende Modell der Grunderfahrungen mit je drei Items für jede latente Variable darf somit höchstens eine manifeste Variable von der Fixierung zwischen den Gruppen gelöst werden.

Die Messinvarianz wurde sowohl für die Gruppe der Mädchen und Jungen als auch zwischen Lernenden mit Informatikunterricht und ohne Informatikunterricht untersucht, um der Fragestellung in c) und d) nachgehen zu können.

Bei der Untersuchung der Messinvarianz bezüglich eines besuchten Informatikunterrichts stellte sich das Problem, dass die Dauer des Informatikunterrichts nicht

eindeutig erhoben werden konnte. Ein weiterer Aspekt ist, dass sich Schülerinnen und Schüler auch im Alltag mit Informatiksystemen auseinandersetzen müssen und dabei Erfahrungen sammeln. Deshalb wurden die Teilnehmerinnen und Teilnehmer in zwei Gruppen aufgeteilt. Eine Gruppe mit mehr als einem Jahr (inf1) und eine Gruppe mit höchstens einem Jahr (info) Informatikunterricht.

Für die Beantwortung der Fragestellungen b) und c), den Ausprägungen der Mittelwerte von GI₁, GI₂ und GI₃ bei Mädchen und Jungen einerseits und dem Ausmaß an einem besuchten Informatikunterricht andererseits, wurde eine zweifaktorielle Varianzanalyse mit anschließendem Post-Hoc-Test mit Bonferroni-Korrektur durchgeführt.

Zusätzlich wurden die latenten Mittelwerte von GI₁, GI₂ und GI₃ mit Hilfe einer konfirmatorischen Mehrgruppen-Faktorenanalyse geschätzt und den Ergebnissen der Post-Hoc-Tests der Varianzanalyse gegenübergestellt (Tabelle 3 und Tabelle 4).

Um die Fragestellung d), der Frage nach der Abhängigkeit von Alter, Geschlecht und Dauer des Informatikunterrichts, beantworten zu können, wurde eine multiple lineare Regression mit den Regressoren Alter, Geschlecht und Informatikunterricht durchgeführt.

2.2 Stichprobe

Eine Gelegenheitsstichprobe von 428 Schülerinnen und Schüler ($n_w = 219$, $n_m = 209$) im Alter von 11 bis 18 Jahren ($M = 14.49$, $SD = 1.47$), wurde im Zeitraum von 2016 bis 2018 an drei Gymnasien und einer Neuen Mittelschule, mit den Angaben zu Geschlecht und der Teilnahme an einem mehr als einjährigen Informatikunterricht, erhoben. Insgesamt wurden 628 Fragebögen retourniert welche sich durch die Beschränkung des zu untersuchenden Alters auf 18 Jahre und der Entfernung von Ausreißern, bestimmt über die Mahalanobis-Distanz, reduzierte. Nicht vollständig ausgefüllte Fragebögen wurden ebenfalls entfernt.

Es gibt unterschiedliche Kriterien für die Wahl der Stichprobengröße. Ein Kriterium ist das Verhältnis von Fällen N zu den zu schätzenden Parametern p , welches mit $N:p$ von 5 bis 10 angegeben wird (Bentler & Chou, 1987; Schreiber et al., 2006). Bei einem Modell mit drei latenten Variablen und je drei manifesten Variablen sind mit den Intercepts 30 Variablen zu schätzen, woraus sich bei gegebener Stichprobengröße pro Gruppe von rund $N = 210$ ein $N:p$ von 7 ergibt. Bei kleinen Stichprobengrößen bleibt zwar der Maximum Likelihood Schätzer (ML) konsistent, aber die Konvergenz wird schlechter und der χ^2 -Test weist korrekt spezifizierte Modelle eher zurück (Bastardo & Antonakis, 2014).

2.3 Erhebungsinstrument

Um die Grunderfahrungen des Informatikunterrichts bei Schülerinnen und Schülern messen zu können, wurde das Inventar zu den Grunderfahrungen des Informatikunterrichts IGI verwendet (Kiener, 2019). Das Messinstrument umfasst drei Subskalen GI₁, GI₂ und GI₃ mit jeweils drei Items, welche mit einer endpunktbeschrifteten fünfstufigen Zustimmungsskala, von „stimme überhaupt nicht zu“ bis „stimme voll und ganz zu“, bewertet werden musste. Die Reliabilitäten der Skalen GI₁ ($\rho = .87$), GI₂ ($\rho = .70$) und GI₃ ($\rho = .73$), bestimmt über Raykovs Rho, genügen den Anforderungen für eine reliable Messung, welche mit .70 als hinreichende Bedingung angegeben wird (Nunnally, 1978; Raykov & Marcoulides, 2016).

2.4 Statistische Analysen

Als Schätzer für die konfirmatorisch untersuchten Mehrgruppen-Modelle wurde Maximum Likelihood (ML) und für die Bewertung der Modelle der χ^2 -Test angewendet, welcher die Abweichung der modellgenerierten zur empirischen Kovarianzmatrix beurteilt. Je kleiner der χ^2 -Wert ausfällt, umso besser ist die Modellanpassung. Damit verknüpft ist der p-Wert, welcher nicht kleiner als .05 sein darf, da sonst die Abweichung signifikant werden würde. Allerdings ist der χ^2 -Test zu streng und lehnt häufig gut fittende Modelle ab (Kline, 2011). Deswegen wird empfohlen, zusätzlich den CFI, RMSEA und SRMR als Fit-Indizes heranzuziehen (Chen, 2007).

Der CFI (Comparative Fit Index) vergleicht den Fit eines Modells mit dem Fit des best fittenden Modells. Der CFI hängt nur geringfügig von der Stichprobengröße ab und bewertet den Fit eines Modells umso besser, je näher er 1.0 ist und sollte für einen guten Fit größer 0.97 sein, wobei der Grenzwert zu nicht fittenden Modellen mit 0.95 angegeben wird (Hu & Bentler, 1999; Sivo et al., 2006). Entscheidend bei der Anwendung des CFI ist aber die Änderung, da sich der CFI bei jeder Modellrestriktion vom konfiguralen Modell bis zum skalaren Modell verschlechtert. Empfohlen wird ein $\Delta\text{CFI} \leq -0.01$ bis zu einem $\Delta\text{CFI} \leq -0.005$ für $n \leq 300$ (Cheung & Rensvold, 2002). Der RMSEA (Root Mean Square Error of Approximation) ist vom Typ ein absoluter Fitindex. Der RMSEA misst die Diskrepanz zwischen der beobachteten Kovarianzmatrix und der modell-implizierten Kovarianzmatrix pro Freiheitsgrad (Steiger, 2000). Der RMSEA hat einen Punktschätzer und einen Intervallschätzer. Ein RMSEA von 0 bedeutet, dass das Modell die Kovarianzmatrix exakt wiedergibt. Ein Wert unter 0.05 wird als gute Modellanpassung betrachtet. Ebenso sollte für einen guten Modellfit das Konfidenzintervall (CI) des Intervallschätzers 0 enthalten. Was bedeutet, dass die linke Intervallgrenze 0.0 sein sollte (Schermelleh-Engel, Moosbrugger & Müller, 2003). Ebenso wie bei dem CFI wird die Änderung des RMSEA zur Beurteilung von Invarianzmodellen herangezogen. Bei Werten von $\Delta\text{RMSEA} > .005$ für $n \leq 300$ ist keine Invarianz mehr indiziert (Chen, 2007). Der

SRMR (Standardized Root Mean Square Residual) ist der Mittelwert aller Differenzen zwischen den datenbasierten und den modellimplizierten Korrelationen. Ein SRMR Wert von 0 zeigt eine perfekte Modellierung an. (Weston & Gore, 2006) Modelle mit SRMR-Werten kleiner als 0.05 werden mit einem guten Fit bewertet und SRMR-Werte zwischen 0.05 und 0.1 als akzeptabler Fit (Schermelleh-Engel et al., 2003). Die sonst noch empfohlenen Indizes AIC und BIC stellen in dieser Untersuchung kein Kriterium für die Beurteilung der Modelle dar, weil sie für genestete Modelle keine Anwendung finden. Der χ^2 -Test ist ausreichend und eignet sich für den Vergleich von genesteten Modellen (Hancock & Mueller, 2013).

Die Analysen zur Messinvarianz wurden in R mit dem Softwarepaket Lavaan 0.6-5, die Varianzanalysen und die Regressionsanalyse mit Jamovi 1.1.7.0 durchgeführt.

3. Ergebnisse

3.1 Invarianzuntersuchung zwischen Mädchen und Jungen

Wesentlich für die Durchführung einer konfirmatorischen Faktorenanalyse und Analyse von Strukturgleichungsmodellen mit ML als Schätzer ist die Voraussetzung der multivariaten Normalverteilung der manifesten Variablen. Der Mardia-Test auf Schiefe und Kurtosis weist eine signifikante Abweichung zur Normalverteilung bezüglich der multivariaten Schiefe auf (Skewness = 372.2, $p < .001$, Kurtosis = 0.31, $p = .76$). Allerdings beträgt bei keinem Item die Schiefe mehr als $|0.8|$, woraus auf eine moderate Verletzung der Normalverteilung geschlossen werden kann. Wegen der Verletzung der Voraussetzung ist es erforderlich, die p -Werte des χ^2 -Tests mit Bootstrap nach Bollen und Stine zu ermitteln (Bühner, 2011).

Die Ergebnisse der Invarianzuntersuchung des Inventars der Grunderfahrungen des Informatikunterrichts zwischen Mädchen und Jungen sind in Tabelle 1 dargestellt. In Modell M₁, dem konfigural invarianten Modell, werden Ladungen, Intercepts und Varianzen in beiden Gruppen unterschiedlich geschätzt, mit dem Ergebnis, dass keine signifikante Verschlechterung der errechneten Varianz-Kovarianzmatrix zu der auf Messwerten basierten Matrix nachgewiesen werden kann ($\chi^2(48) = 64.78$, $p = .083$, RSMEA = .04, CI [.00–.06], SRMR = .046, CFI = .986). Ebenso wenig verschlechtert die Restriktion der Ladungen in Modell M₂ die Modellanpassung, womit metrische Invarianz angenommen werden kann ($\chi^2(54) = 72.03$, $p = .073$, RSMEA = .039, CI [.00–.06], SRMR = .053, CFI = .984, $\Delta CFI = -.001$, $\Delta RMSEA = -.001$, $p_{diff} = .298$).

Erst bei der Gleichsetzung der Intercepts im skalaeren Modell M₃ zeigen die absoluten Fitindizes $\chi^2(60) = 89.19$, $p = .023$, RSMEA = .048, CI [.03–.06] und die inkrementellen Indizes, $\Delta CFI = .010$, $\Delta RMSEA = +.009$, $p_{diff} = .009$, eine signifikante Verschlechterung, womit nicht von einer skalaren Invarianz ausgegangen werden kann.

TAB. 1. Indizes der Invarianz-Modelle zwischen Mädchen ($n = 219$) und Jungen ($n = 209$)

Modell	χ^2 (df)	p_{χ^2}	RSMEA [90% CI]	SRMR	CFI	inkrementeller Vergleich			
						Referenz	Δ CFI	Δ RMSEA	p_{diff}
M1 con- figural	64.78 (48)	.083	.040 [.00 - .06]	.046	.986				
M2 metric	72.03 (54)	.073	.039 [.00 - .06]	.053	.984	M1	-0.001	-0.001	.298
M3 scalar	89.19 (60)	.023	.048 [.03 - .07]	.057	.975	M2	-0.010	+0.009	.009
M4 part scalar	80.16 (58)	.050	.041 [.00 - .06]	.054	.982	M2	-0.004	+0.007	.149
M5 part mean	169.8 (62)	.001	.090 [.07 - .11]	.132	.970	M4	-0.012	+0.049	<.001
Akzept. Kriterien		>.05	<.05 [.00 - ...]	<.10	>0.97		<-.005	<+.005	>.05

Anmerkungen: $p_{\chi^2} = p$ -Wert des χ^2 -Tests (Bollen-Stine Bootstrap), $p_{diff} = p$ -Wert des χ^2 -Differenzen Tests (in R ANOVA genannt), Werte außerhalb der annehmbaren Kriterien sind fett dargestellt.

Jene Items, welche die skalare Modellanpassung deutlich verschlechtern, wurden über die Differenz der vom Modell berechneten Mittelwerte zu den empirischen Mittelwerten der Items gefunden. Die größte Abweichung zeigt sich bei dem Item der Subskala GI₃, „Fehlermeldungen helfen mir im Umgang mit Informatiksystemen“ (Mädchen +0.123, Jungen -0.138), was bedeutet, dass Mädchen den Beitrag von Fehlermeldungen zu den Grunderfahrungen GI₁ stärker als Jungen bewerten.

Im partiell skalaren Modell M₄ wurde die Restriktion der Gleichheit der Achsenabschnitte für das Item „Fehlermeldungen helfen mir im Umgang mit Informatiksystemen“ in beiden Gruppen aufgehoben. Dadurch wurde, sowohl von den absoluten Indizes als auch von den inkrementellen Indizes, vom Δ RMSEA = +.007 ausgenommen, ein akzeptabler Modellfit erreicht (Tabelle 1).

Mit dem Gleichsetzen der latenten Mittelwerte von GI₁, GI₂ und GI₃ zwischen Mädchen und Jungen in Modell M₅, verschlechterten sich alle Indizes, woraus geschlossen werden kann, dass sich die Mittelwerte der Skalen der Grunderfahrungen des Informatikunterrichts zwischen Schülerinnen und Schülern unterscheiden ($\chi^2(62) = 169.8$, $p = <.001$, RSMEA = .09, CI [.07–.11], SRMR = .132, CFI = .9700, Δ CFI = -.012, Δ RMSEA = +.049, $p_{diff} = <.001$) (Tabelle 1).

3.2 Invarianzuntersuchung zum Ausmaß des Informatikunterrichts

Die Ergebnisse der schrittweisen Restriktionen sind in Tabelle 2 dargestellt. Die Indizes des Modells MI₁ weisen auf eine konfigurale Invarianz hin ($\chi^2(48) = 63.61$,

$p = .091$, RSMEA = .039, CI [.00–.06], SRMR = .054, CFI = .9878). Die Gleichsetzung der Ladungen im metrischen Modell MI2 verschlechtert die Modellanpassung, welche durch die absolute Änderung zum saturierten Modell $\chi^2(54) = 78.34$, $p = .031$, RSMEA = .047, CI [.02–.07], SRMR = .059, CFI = .9809 als auch durch die Änderungsindizes $\Delta\text{CFI} = -.007$, $\Delta\text{RMSEA} = +.008$, $p_{\text{diff}} < .023$ signifikant ist.

Die Modifikationsindizes weisen auf das Item „Fehlermeldungen helfen mir im Umgang mit Informatiksystemen“ hin, welches das χ^2 am meisten verringern würde. Geprüft wurde daraufhin das partial metrische Modell MI3, mit für dieses Item in beiden Gruppen unterschiedlich geschätzten Ladungen. Dieses Modell weist in allen Indizes auf einen guten Modellfit hin (Tabelle 2).

Da schon im metrischen Modell die Ladungen eines Items freigesetzt wurden, müssen auch für die Intercepts dieses Items die Restriktionen der Gleichheit aufgehoben werden, um zum partial skalaren Modell MI4 zu gelangen. Anhand der Fitindizes $\chi^2(58) = 71.96$, $p = 1.03$, RSMEA = .034, CI [.00–.06], SRMR = .056, CFI = .9891 und den inkrementellen Indizes $\Delta\text{CFI} = -.000$, $\Delta\text{RMSEA} = -.001$, $p_{\text{diff}} = 1.0$ ist zu schließen, dass für die Skalen partiell skalare Invarianz angenommen werden kann.

Durch das Gleichsetzen der latenten Mittelwerte im Modell MI5 verschlechtert sich der globale Modellfit $\chi^2(61) = 98.66$, $p = .005$, RSMEA = .054 CI [.03–.07] und auch alle inkrementellen Indizes weisen auf einen Missfit hin $\Delta\text{CFI} = -.020$, $\Delta\text{RMSEA} = +.022$, $p_{\text{diff}} < .001$. Daraus lässt sich schließen, dass die latenten Mittelwerte für beide Gruppen unterschiedlich sind (Tabelle 2).

TAB. 2. Indizes der Invarianz-Modelle zwischen den Gruppen mit mehr als einem Jahr Informatikunterricht ($n = 214$) und höchstens einem Jahr Informatikunterricht ($n = 214$)

Modell	χ^2 (df)	p_{χ^2}	RSMEA [90% CI]	SRMR	CFI	inkrementeller Vergleich			
						Referenz	ΔCFI	ΔRMSEA	p_{diff}
MI1 config- ural	63.61 (48)	.091	.039 [.00 - .06]	.054	.9878				
MI2 metric	78.34 (54)	.031	.047 [.02 - .07]	.059	.9809	MI1	-.007	+.008	.023
MI3 part metric	66.90 (53)	.130	.035 [.00 - .06]	.055	.9891	MI1	-.001	-.004	.130
MI4 part scalar	71.96 (58)	.103	.034 [.00 - .06]	.056	.9891	MI3	.000	-0.01	1.00
MI5 part mean	98.66 (61)	.005	.054 [.03 - .07]	.079	.9705	MI4	-.020	+.022	<.001
Akzept. Kriterien		>.05	<.05 [.00 - ...]	<.01	>0.97		<-.005	<+.005	>.05

Anmerkungen: $p_{\chi^2} = p$ -Wert des χ^2 -Tests (Bollen-Stine Bootstrap), $p_{\text{diff}} = p$ -Wert des χ^2 -Differenzen Tests (in R ANOVA genannt), Werte außerhalb der annehmbaren Kriterien sind fett dargestellt.

3.3 Untersuchung der latenten Mittelwerte

Die Invarianzuntersuchungen haben ergeben, dass für das Messinstrument zu den Grunderfahrungen des Informatikunterrichts auf Grund des Prüfverfahrens partial skalare Invarianz, mit nur einer frei gestellten manifesten Variablen, angenommen werden kann und daher Mittelwertvergleiche möglich sind (Putnick & Bornstein, 2016).

Es wurde für jede der drei Skalen der Grunderfahrungen des Informatikunterrichts je eine zweifaktorielle Varianzanalyse mit den Gruppen Gender (m, w) und Informatikunterricht (inf 0, inf 1) mit einem anschließendem Post-Hoc-Test mit Bonferroni-Korrektur der p-Werte durchgeführt. Der Test auf Varianzhomogenität nach Levene, eine Voraussetzung zur Durchführung der Varianzanalyse, ist bei allen drei Skalen nicht signifikant und es kann eine Varianzanalyse durchgeführt werden (Gl1: $p = .374$, Gl2: $p = .217$, Gl3: $p = 0.762$).

Die Berechnung der zweifaktoriellen ANOVA für Gl1 ergab sowohl einen signifikanten Haupteffekt zwischen Mädchen und Jungen ($F(1,424) = 69.74$, $p < .001$, $\eta_p^2 = .141$, $r = .38$) als auch dafür, ob mehr als ein Jahr Informatikunterricht besucht wurde ($F(1,424) = 16.42$, $p < .001$, $\eta_p^2 = .037$, $r = .19$). Den Daten ist anhand der relativen Effektstärke r zu entnehmen, dass zwischen Jungen und Mädchen bei der Grunderfahrung eins des Informatikunterricht ein mittlerer Effekt ($r = .38$) und bei Lernenden mit mehr als einem Jahr Informatikunterricht ein geringer Effekt ($r = .19$), beobachtet werden konnte. Bei relativen Effektstärken $r > 0.5$, welches Cohens $d > 1.2$ entspricht, kann von großen Effekten ausgegangen werden (Cohen, 1988; Field, 2011).

Für Gl2 sind ebenfalls die Haupteffekte des Faktors Gender $F(1,424) = 42.63$, $p < .001$, $\eta_p^2 = .091$, $r = .3$ und der Faktor Informatikunterricht mit den Werten $F(1,424) = 7.31$, $p = .007$, $\eta_p^2 = .017$, $r = .13$ signifikant. Die Effekte sind allerdings, gemessen an den relativen Effektstärken $r = .3$ und $r = .13$, gering.

Bei Gl3 zeigt sich ein anderes Bild. Hier wird kein signifikanter Unterschied weder zwischen den Geschlechtern $F(1,424) = 2.162$, $p = .142$ noch bezüglich des Faktors, ob mehr als ein Jahr Informatikunterricht besucht wurde $F(1,424) = 0.279$, $p = .918$, angezeigt.

Die Interaktionseffekte zwischen Mädchen und Jungen (Gender) und Schülerinnen und Schülern mit mehr als einem Jahr Informatikunterricht und höchstens einem Jahr Informatikunterricht (Informatikunterricht) zeigten sich nicht signifikant bezüglich der Skalen der Grunderfahrungen des Informatikunterrichts, Gl1 mit $F(1,424) = 0.745$, $p = .389$, Gl2 mit $F(1,424) = .096$, $p = 0.57$ und Gl3 mit $F(1, 424) = 0.0011$, $p = .918$.

In Tabelle 3 sind die Mittelwerte und Änderungen zwischen den Gruppen m (männlich) und w (weiblich) und in Tabelle 4 zwischen den Gruppen inf1 (mehr als ein Jahr Informatikunterricht) und info (höchstens ein Jahr Informatikunterricht)

dargestellt. Die empirischen Mittelwerte sind denen, mit dem über das partial skalare Modell M4 errechneten latenten Mittelwerten, gegenübergestellt.

TAB. 3. Mittelwerte und Änderungsmaße von Mädchen ($n_w = 219$) und Jungen ($n_m = 209$).

Skala	Empirisch					Modellbasiert aus M4			
	M_m (SD)	M_w (SD)	ΔM	t(df)	p_b	M_m (SE)	M_w (SE)	ΔM	p_{χ^2}
GI1	3.287 (0.953)	2.525 (0.880)	.762	8.35 (424)	<.001	3.411 (0.071)	2.670 (0.068)	.741	<.001
GI2	4.314 (0.592)	3.904 (0.661)	.410	6.53 (424)	<.001	4.483 (0.047)	4.155 (0.059)	.328	<.001
GI3	3.703 (0.840)	3.578 (0.853)	.125	1.47 (424)	.142	3.464 (0.078)	3.530 (0.072)	.066	.570

Anmerkungen: t = t-Wert des Post-Hoc Tests, df = Freiheitsgrad, p_b = p-Wert mit Bonferro-ni-Korrektur, p_{χ^2} = p-Wert des χ^2 -Differenzen Tests, m = männlich, w = weiblich

Schüler antworten mit einem signifikant ($p_b < .001$ und $p_{\chi^2} < .001$) höheren Wert auf der Zustimmungsskala für GI1 ($\Delta M = .762$, $t(424) = 8.35$) und für GI2 ($\Delta M = .41$, $t(424) = 6.53$) als Schülerinnen. Für die Skala GI3 wird kein signifikanter Unterschied ($p_b = .128$ und $p_{\chi^2} = .57$) zwischen Mädchen und Jungen angezeigt (Tabelle 3).

TAB. 4. Mittelwerte und Änderungsmaße von Lernenden mit mehr als einem Jahr Informatik-unterricht ($n_{inf1} = 214$) und höchstens einem Jahr Informatikunterricht ($n_{info} = 214$).

Skala	Empirisch					Modellbasiert aus MI4			
	M_{inf1} (SD)	M_{info} (SD)	ΔM	t(df)	p_b	M_{inf1} (SE)	M_{info} (SE)	ΔM	p_{χ^2}
GI1	3.106 (1.017)	2.688 (0.922)	.418	4.05 (424)	<.001	3.224 (0.081)	2.841 (0.084)	.383	<.001
GI2	4.204 (0.620)	4.005 (0.684)	.199	2.70 (424)	.007	4.379 (0.081)	4.252 (0.086)	.126	.125
GI3	3.667 (0.887)	3.612 (0.809)	.055	0.53 (424)	.597	3.514 (0.085)	3.481 (0.086)	.003	.756

Anmerkungen: t = t-Wert des Post-Hoc-Tests, df = Freiheitsgrad, p_b = p-Wert mit Bonferro-ni-Korrektur, p_{χ^2} = p-Wert des χ^2 -Differenzen Tests, $inf1$ = mehr als ein Jahr Informatikunter-richt, $info$ = höchstens ein Jahr Informatikunterricht

Sowohl die Post-Hoc-Tests als auch das partial skalare Modell MI4 weisen für GI1 einen signifikanten ($p_b < .001$, $p_{\chi^2} < .001$) Unterschied ($\Delta M = .383$, $t(424) = 4.05$) bei Lernenden mit mehr als einem Jahr Informatikunterricht ($inf1$) im Vergleich zu je-nen mit höchstens einem Jahr Informatikunterricht ($info$) auf. Für GI2 errechnet das partial skalare Modell IM4 nicht signifikante Mittelwertsunterschiede ($\Delta M = .126$, $p_{\chi^2} = .1235$), sehr wohl aber der Post-Hoc-Test ($\Delta M = .199$, $t(424) = 2.7$, $p_b = .007$). Kei-

ne signifikanten Auswirkungen des Ausmaßes vom Informatikunterricht zwischen Jungen und Mädchen können für Gl₃ angenommen werden (Tabelle 4).

3.4 Regressionsanalyse

Zur Verifikation der Zusammenhangshypothese, dass die Grunderfahrungen des Informatikunterrichts mit dem Alter zunehmen, wurde eine multiple Regressionsanalyse mit den Regressoren Alter, Geschlecht und Informatikunterricht durchgeführt. Die Subskalen Gl₁, Gl₂ und Gl₃ wurden getrennt voneinander untersucht. In dieser Analyse wird zu den schon zuvor untersuchten Abhängigkeiten noch das Alter mit einbezogen.

Eine Durbin-Watson d-Analyse wurde durchgeführt, um die Unabhängigkeit der Residuen, eine Voraussetzung der multiplen linearen Regression, zu überprüfen. Die Analyse ergab für Gl₁ (DW = 1.7, $p = .001$), für Gl₂ (DW = 1.6, $p < .001$) und für Gl₃ (DW = 1.9, $p = .08$). DW-Werte um 2, das heißt zwischen 1.5 und 2.5, werden als akzeptabel angesehen (Field, 2011).

Eine weitere Voraussetzung für eine multiple lineare Regression ist die Unabhängigkeit der Prädiktoren, welche über den Varianzinflationsfaktor (VIF) geprüft wurde. Für die Subskalen Gl₁, Gl₂ und Gl₃ haben die Regressoren Alter, Geschlecht und Informatik einen VIF-Wert von jeweils 1.05, 1.01 und 1.05. VIF-Werte über 10 lassen auf eine Multikollinearität schließen (Field, 2011).

Die Grunderfahrungen Gl₁ sind nicht signifikant vom Alter abhängig. Lediglich die zuvor schon untersuchten signifikanten Abhängigkeiten vom Geschlecht und dem Ausmaß des Informatikunterrichts bestätigen sich, zu entnehmen aus der Tabelle 5. Gl₁ kann formal ausgedrückt werden durch $Gl_1 = 3.71 - 0.072 * \text{Gender} + 0.162 * \text{Informatikunterricht}$. Die Variable Gender ist mit „männlich“ = 0 und „weiblich“ = 1 und die Variable Informatikunterricht mit „höchstens ein Jahr Informatikunterricht“ = 0 sowie mit „mehr als ein Jahr Informatikunterricht“ = 1 einzusetzen.

Gl₂ hängt signifikant vom Geschlecht und vom Ausmaß des Informatikunterrichts, aber nicht vom Alter ab. Der formale Zusammenhang ist $Gl_2 = 4.18 - 0.396 * \text{Gender} + 0.162 * \text{Informatikunterricht}$, wobei die Variablen Gender und Informatikunterricht wie zuvor kodiert sind.

Anders als zuvor bei den latenten Variablen Gl₁ und Gl₂, bei denen keine signifikante Abhängigkeit vom Alter angenommen werden kann, zeigt Gl₃ ausschließlich eine signifikante Abhängigkeit vom Lebensalter mit einer sehr geringen Aufklärung der Varianz von 3.4 % (Tabelle 5). Die Altersabhängigkeit wird formal ausgedrückt durch $Gl_3 = 5.09 - 0.10 * \text{Alter}$, das Alter ist in Jahren einzusetzen.

TAB. 5. Regressionsmodelle der Subskalen der Grunderfahrungen des Informatikunterrichts ($n = 428$)

Skala	Prädiktor	b_0	b	SE	β	t	p_t	R^2
GI1		3.71		.44		8.53	<.001	.183
	Alter Jahre		-.044	.03	-.065	-1.44	.149	
	Gender m -> w		-.072	.09	-.037	-8.26	<.001	
	Inf 0 -> 1		.38	.09	.191	4.25	<.001	
GI2		4.18		.302		13.82	<.001	.112
	Alter Jahre		0.004	.021	.008	0.17	.867	
	Gender m -> w		-0.396	.061	-.300	-6.53	<.001	
	Inf 0 -> 1		0.162	.062	.123	2.62	.009	
GI3		5.09		.405		12.56	<.001	.034
	Alter Jahre		-0.100	.028	-.172	-3.53	<.001	
	Gender m -> w		-0.103	.081	-.061	-1.26	.208	
	Inf 0 -> 1		0.105	.083	.062	1.26	.208	

Anmerkungen: GI1 $F(3, 424) = 31.7$ $p < .001$, GI2 $F(3, 424) = 17.85$, $p < .0001$, GI3 $F(3, 424) = 5.00$, $p = .002$, b_0 = Achsenabschnitt, b = Regressionskonstante, β = standardisierte Regressionskonstante, SE = Standardfehler, p_t = p -Wert der t -Statistik, Gender männlich $m = 0$, weiblich $w = 1$, Inf = 0 höchstens ein Jahr Informatikunterricht, Inf = 1 mehr als ein Jahr Informatikunterricht

4. Diskussion

In dieser Studie konnte anhand einer Gelegenheitsstichprobe von 428 Schülerinnen und Schülern für das Inventar zu den Grunderfahrungen des Informatikunterrichts zumindest eine partielle skalare Messinvarianz nachgewiesen werden.

Der IGI erfüllt zwischen Mädchen und Jungen sowohl konfigurale als auch metrische Invarianz, aber nur partiell die skalare Messinvarianz, da das Item „Fehlermeldungen helfen mir im Umgang mit Informatiksystemen“ durch die Gleichsetzung der Intercepts in beiden Gruppen mit einer signifikanten Verschlechterung der modellbasierten Kovarianzmatrix einhergeht.

Zwischen Lernenden mit mehr als einem Jahr und jenen mit höchstens einem Jahr Informatikunterricht wurde ebenfalls eine konfigurale Messinvarianz nachgewiesen. Eine metrische Messinvarianz zwischen diesen Gruppen wurde wiederum durch das Item „Fehlermeldungen helfen mir im Umgang mit Informatiksystemen“

verhindert. Eine partielle metrische Invarianz und in der Folge auch eine nur partielle skalare Invarianz konnte durch das Freisetzen der Ladungen und der Intercepts für dieses Item, über die Gruppen hinweg, erreicht werden.

Ein Inventar mit Messinvarianz ist gut geeignet, die kontinuierliche Entwicklung von Lernenden zu Inhalten von fundamentalen Ideen, wie sie nach Schwill (1993) definiert werden, zu messen. Die Verletzung der Messinvarianz durch ein Item weist auf eine Änderung der Sichtweise hin (van de Schoot et al., 2012). Dies kann ein Hinweis auf die Auflösung eines Misskonzeptes, ein aufgrund von Vorerfahrungen entwickeltes fehlerhaftes mentales Modell sein. Eine für die Informatikdidaktik sehr breit angesetzte Definition fasst „misconceptions“, „incorrect understandings“, „difficulties“, „mistakes“, und „bugs“ zu Misskonzepten zusammen (Sorva, 2013). Entsprechend dieser Sichtweise kann das Item „Fehlermeldungen helfen mir im Umgang mit Informatiksystemen“ unter Misskonzepte subsumiert werden, welche in der Informatikdidaktik für Lehrpersonen als Herausforderung gelten sie aufzulösen (Eckerdal et al., 2006; Qian & Lehman, 2017).

Somit konnte in dieser Studie gezeigt werden, dass durch einen längeren Informatikunterricht Schülerinnen und Schüler ihr mentales Modell revidieren, um Fehlermeldungen, verursacht durch Informatiksysteme, handhaben zu können. Das gilt sowohl für Jungen als auch für Mädchen, obwohl Mädchen trotz des schon höheren Intercepts des Items dennoch über die Ladung des Items auf die latente Variable G₃ einen Entwicklungssprung zeigen.

Die Unterschiede der latenten Mittelwerte von G₁ weisen im Vergleich zwischen Mädchen und Jungen einen mittleren Effekt aus. Hingegen macht sich nur ein kleiner Effekt zwischen Lernenden mit mehr als einem Jahr dauernden Informatikunterricht gegenüber der Vergleichsgruppe mit höchstens einem Jahr Informatikunterricht bemerkbar. Da die erste Grunderfahrung des Informatikunterrichts die Wahrnehmung der eigenen Kompetenz widerspiegelt, lässt sich aus den Ergebnissen folgern, dass Jungen die eigene Kompetenz bezüglich Informatiksystemen stärker einschätzen als Mädchen, wobei beide Geschlechter lediglich unterhalb des neutralen Bereichs auf der Zustimmungsskala antworten (Kiener, 2019).

Das bestätigt die in anderen Studien für die zu G₁ verwandten Variablen gefundenen Ergebnisse, wie bei Magenheim und Schulte (2005) für das „Selbstvertrauen im Umgang mit Computern“ oder auch bei Lishinski, Yadav, Good und Enbody, (2016) für „computer related self-efficacy“ im Erlernen von Programmieren und „subjective technical competence“ bei Brauner, Leonhardt, Ziefle und Schroeder (2010). In den zuvor genannten Studien wurde gezeigt, dass dies entscheidend für die weitere Entwicklung bezüglich informatischer Kompetenzen ist, deshalb sollte im Informatikunterricht die Förderung der ersten Grunderfahrung des Informatikunterrichts an Bedeutung gewinnen.

Lernende mit wenig oder keinem Informatikunterricht schätzen ihre Kompetenz nur geringfügig kleiner im Vergleich zu jenen mit ausreichend Informatikun-

terrichtet ein, da auch diese Gruppe Computererfahrung sammelt. Dieses Ergebnis stimmt mit einer Untersuchung der „subjective technical competence“ überein, welche allein mit der Dauer der Computernutzung zunimmt (Brauner et al., 2010).

Beim Erkennen der hinter den Informatiksystemen stehenden Algorithmen, welches der zweiten Grunderfahrung des Informatikunterrichts entspricht, lässt sich für Jungen wiederum ein höherer Wert im Vergleich zu Mädchen nachweisen, allerdings mit einer geringen Effektstärke. Das gleiche Bild lässt sich zwischen der Gruppe mit mehr als einem Jahr und jener mit höchstens einem Jahr Informatikunterricht erkennen, wobei die erste Gruppe geringfügig höhere Werte auf der Zustimmungsskala aufweist. Alle Gruppen antworten mit hoher Zustimmung auf der fünfteiligen Likert-Skala, was für den Informatikunterricht bedeutet, dass Schülerinnen und Schüler schon mit Vorerfahrungen in den Informatikunterricht kommen. Dies stellt wiederum eine Voraussetzung für algorithmisches Denken dar, welches als Schlüssel für das Verständnis von Informatiksystemen gilt (Futschek, 2006).

Die dritte Grunderfahrung des Informatikunterrichts betrachtet die Einstellung zur Anwendbarkeit des Erlernten. Dabei überrascht, dass lediglich das Alter eine signifikante Abnahme, bei einem Altersunterschied von zehn Jahren um eine ganze Einheit auf der fünfteiligen Zustimmungsskala, bewirkt. Die negative Regression mit dem Alter bedeutet aber nicht, dass GI₃ mit dem Alter abnimmt, welches nur für eine Längsschnittstudie gültig wäre. Es lässt sich aus den erhobenen Daten weder ein signifikanter Unterschied zwischen Jungen und Mädchen noch eine signifikante Auswirkung eines Informatikunterrichts nachweisen. Die Mittelwerte liegen für alle Gruppen im oberen neutralen Bereich um 3,6 und lassen damit für den Informatikunterricht noch ein weites Feld offen, um Lernende von der Anwendbarkeit ihrer Fähigkeiten zu überzeugen. Wie wichtig GI₃ für die spätere Studien- und Berufswahl ist, zeigen Bieg und Röddiger (2017) in einer Untersuchung über die Wahlintention zu MINT-Studienfächern und MINT-Berufen. Dabei korrelierte die Subskala „Nützlichkeit“ für alle Mintfächer mit $r > 0.6$ mit der Wahlintention für ein Fach sehr hoch. Außer für Mathematik konnte in allen anderen MINT-Fächern ebenfalls kein signifikanter Gender-Gap für „Nützlichkeit“ nachgewiesen werden.

5. Limitierungen

Obwohl keine der untersuchten Gruppen überrepräsentiert ist, sind die Ergebnisse aufgrund der gezogenen Gelegenheitsstichprobe nicht repräsentativ. Die Ergebnisse der Untersuchung bei den latenten Mittelwerten der Skalen des IGI zwischen den Gruppen mit mehr oder weniger Informatikunterricht lassen keine Aussage zu, wie wirksam ein Informatikunterricht bezüglich der Ausprägungen der Mittelwerte ist. Um einen Effekt eines Informatikunterrichts bestimmen zu können, sollten die Variable der außerschulischen Nutzung und die dadurch erlangten Erfahrungen in einem Quasiexperiment kontrolliert werden.

Literatur

- Bastardo, N. & Antonakis, J. (2014). Sample Size Requirement for Unbiased Estimation of Structural Equation Models: A Monte Carlo Study. *Academy of Management Proceedings*, 2014(1), 13405.
- Bentler, P. M. & Chou, C. P. (1987). Practical Issues in Structural Modeling. *Sociological Methods & Research*, 16(1), 78–117.
- Bieg, M. & Röddiger, A. P. (2017). Befragung zur MINT-Studienfach- und -Berufswahl von Abiturientinnen und Abiturienten in Baden-Württemberg. Bieg, M., & Röddiger, A. P. (2017). Befragung zur MINT-Studienfach- und -Berufswahl von Abiturientinnen und Abiturienten in Baden-Württemberg. *Lehren & Lernen: Zeitschrift für Schule und Innovation in Baden-Württemberg* (7), 24–32.
- Brauner, P., Leonhardt, T., Ziefle, M. & Schroeder, U. (2010). The Effect of Tangible Artifacts, Gender and Subjective Technical Competence on Teaching Programming to Seventh Graders. In J. Hromkovič, R. Královič & J. Vahrenhold (Hrsg.), *Teaching fundamental concepts of informatics. 4th International Conference on Informatics in Secondary Schools - Evolution and Perspectives, ISSEP 2010, Zurich, Switzerland, January 13–15, 2010; proceedings* (Lecture notes in computer science, Bd. 5941, Bd. 5941, S. 61–71). Berlin: Springer.
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (PS Psychologie, 3., aktualisierte und erw. Aufl.). München: Pearson Studium.
- Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural equation modeling: a multidisciplinary journal*, 14(3), 464–504.
- Cheung, G. W. & Rensvold, R. B. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural equation modeling: a multidisciplinary journal*, 9(2), 233–255.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2. ed.). Hillsdale, NJ: Erlbaum.
- Eckerdal, A., McCartney, R., Moström, J. E., Ratcliffe, M., Sanders, K. & Zander, C. (2006). Putting threshold concepts into context in computer science education. *ACM SIGCSE Bulletin*, 38(3), 103.
- Eickelmann, B., Bos, W., Gerick, J., Goldhammer, F., Schaumburg, H., Schwippert, K. et al. (2019). *ICILS 2018 #Deutschland. Computer- und informationsbezogene Kompetenzen von Schülerinnen und Schülern im zweiten internationalen Vergleich und Kompetenzen im Bereich Computational Thinking* (1. Auflage). Münster: Waxmann.
- Field, A. (2011). *Discovering statistics using SPSS. (and sex and drugs and rock 'n' roll)* (3. ed., reprinted.). Los Angeles, Calif.: Sage.

- Futschek, G. (2006). Algorithmic Thinking: The Key for Understanding Computer Science. In R. T. Mittermeir (Hrsg.), *Informatics education - the bridge between using and understanding computers. International Conference on Informatics in Secondary Schools - Evolution and Perspectives, ISSEP 2006, Vilnius, Lithuania, November 7-11, 2006; proceedings* (Lecture notes in computer science, Bd. 4226, S. 159–168). Berlin: Springer.
- Hancock, G. R. & Mueller, R. O. (2013). *Structural equation modeling. A second course (Quantitative methods in education and the behavioral sciences)*. Charlotte, NC: Information Age Publishing Inc.
- Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, 6(1), 1–55.
- Kiener, A. (2019). Entwicklung eines Messinstruments zu den Grunderfahrungen des Informatikunterrichts. *Teaching Mathematics and Computer Science*, 17(2), 159–178.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling (Methodology in the social sciences, 3. ed.)*. New York: Guilford Press.
- Lishinski, A., Yadav, A., Good, J. & Enbody, R. (2016). *Learning to program. In ICER'16. Proceedings of the 2016 ACM Conference on International Computing Education Research : September 8-12, 2016, Melbourne, Australia* (Bd. 24, S. 211–220). New York, NY: ACM Association for Computing Machinery.
- Loyalka, P., Liu, O. L., Li, G., Chirikov, I., Kardanova, E., Gu, L. et al. (2019). *Computer science skills across China, India, Russia, and the United States. Proceedings of the National Academy of Sciences of the United States of America*, 116(14), 6732–6736.
- Magenheim, J. & Schulte, C. (2005). Erwartungen und Wahlverhalten von Schülerinnen und Schülern gegenüber dem Schulfach Informatik – Ergebnisse einer Umfrage. *INFOS*, 111–122.
- Massek, C., Drossel, K. & Eickelmann, B. (2018). Eine geschlechtsspezifische Analyse zum schulischen Erwerb computerbezogener Kompetenzen in Tablet-Klassen am Gymnasium. *Medienimpulse*, 56(4).
- Milfont, T. & Fischer, R. (2010). Testing measurement invariance across groups: applications in cross-cultural research. *International Journal of Psychological Research*, 3(1), 111–121.
- Nunnally, J. C. (1978). *Psychometric theory* (McGraw-Hill series in psychology, 2. ed.). New York: McGraw-Hill.
- Ponzo, M. (2011). Does the Way in which Students Use Computers Affect their Performance? *Journal of Economic and Social Research*, 13(2), 1–27.
- Putnick, D. L. & Bornstein, M. H. (2016). Measurement Invariance Conventions and Reporting: The State of the Art and Future Directions for Psychological Research. *Developmental review*, 41, 71–90.

- Qian, Y. & Lehman, J. (2017). Students' Misconceptions and Other Difficulties in Introductory Programming. *ACM Transactions on Computing Education*, 18(1), 1–24.
- Raykov, T. & Marcoulides, G. A. (2016). Scale Reliability Evaluation Under Multiple Assumption Violations. *Structural equation modeling: a multidisciplinary journal*, 23(2), 302–313.
- Richter, T., Naumann, J. & Horz, H. (2002). Computer Literacy, computerbezogene Einstellungen und Computernutzung bei männlichen und weiblichen Studierenden. In M. Herczeg (Hrsg.), *Mensch & Computer 2002. Vom interaktiven Werkzeug zu kooperativen Arbeits- und Lernwelten* (Berichte des German Chapter of the ACM, Bd. 56, Bd. 55, S. 71–80). Stuttgart: B.G. Teubner.
- Richter, T., Naumann, J. & Horz, H. (2010). Eine revidierte Fassung des Inventars zur Computerbildung (INCOBI-R). *Zeitschrift für Pädagogische Psychologie*, 24(1), 23–37.
- Schermelleh-Engel, K., Moosbrugger, H. & Müller, H. (2003). Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures. *Methods of psychological research online*, 8(2), 23–74.
- Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A. & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of educational research*, 99(6), 323–338.
- Schwill, A. (1993). Fundamentale Ideen der Informatik. *Zentralblatt für Didaktik der Mathematik*, 25(1), 20–31.
- Sivo, S. A., Fan, X., Witta, E. L. & Willse, J. T. (2006). The Search for "Optimal" Cutoff Properties: Fit Index Criteria in Structural Equation Modeling. *The Journal of Experimental Education*, 74(3), 267–288.
- Sorva, J. (2013). Notional machines and introductory programming education. *ACM Transactions on Computing Education*, 13(2), 1–31.
- Steenkamp, J.-B. E. M. & Baumgartner, H. (1998). Assessing Measurement Invariance in Cross-National Consumer Research. *Journal of Consumer Research*, 25(1), 78–107.
- Steiger, J. H. (2000). Point Estimation, Hypothesis Testing, and Interval Estimation Using the RMSEA: Some Comments and a Reply to Hayduk and Glaser. *Structural equation modeling: a multidisciplinary journal*, 7(2), 149–162.
- Van de Schoot, R., Lugtig, P. & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, 9(4), 486–492.
- Vandenberg, R. J. & Lance, C. E. (2000). A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organizational Research Methods*, 3(1), 4–70.
- Weston, R. & Gore, P. A. (2006). A Brief Guide to Structural Equation Modeling. *The Counseling Psychologist*, 34(5), 719–751.