

Mannheimer Inventar zur Testung der Urteilsakkuratheit – MITU

[Mannheim inventory for testing judgement accuracy – MITU]

Karina Karst, Meike Bonefeld & Jan Siebert

Ziel

Beim Mannheimer Inventar zur Testung der Urteilsakkuratheit (MITU) handelt es sich um ein virtuelles Klassenzimmer. Innerhalb dieser virtuellen Umgebung sollen Urteile durch Lehrkräfte oder angehende Lehrkräfte über Schüler_innen in einer Annäherung an ein reales Klassensetting erfasst werden.

Die Merkmale der in der aktuellen Form des virtuellen Klassenzimmers enthaltenen Schüler_innen (N = 12) sind experimentell variiert, um Effekte verschiedener Schülermerkmale (Leistungsniveau, Geschlecht, Migrationshintergrund) auf die Urteilsbildung und auf die Akkuratheit des Urteils bestimmen zu können.

Daneben können mit dem Inventar auch Merkmale des Urteilers und/ oder Merkmale des Urteils als Determinanten der Urteilsakkuratheit experimentell variiert werden.

Potentielle Erweiterung

Die im Folgenden erläuterten Schülerbeschreibungen können beliebig, um Faktoren von denen anzunehmen ist, dass sie (fälschlicherweise) das Lehrerurteil beeinflussen erweitert werden.

Die Schulklasse [Klassenzimmer]

Die virtuelle Schulklasse besteht aus zwölf Schüler_innen (6 Jungen, 3 davon mit Migrationshintergrund; 6 Mädchen, 2 davon mit Migrationshintergrund, alle variierend in ihrem Leistungsniveau). Die in der virtuellen Schulklasse präsentierten Namen der Schüler_innen sind fiktiv und im Rahmen von Vortests auf verschiedene Merkmale hin pilotiert.

Bei den Informationen, die die Lehrkräfte über die Schüler_innen abseits des Namens erhalten, handelt es sich um reale Testergebnisse aus einem 15-minütigen Test zur Multiplikation, den Schüler_innen der 2. Klasse unmittelbar nach einer videographierten Doppelstunde zur „Einführung in die Multiplikation“ bearbeitet haben. Die Testerhebung sowie die Videographie des Unterrichts waren Teil des größer angelegten Grundschulprojekts PERLE (Persönlichkeits- und Lernentwicklung von Grundschulkindern; Lipowsky, Faust & Kastens, 2013).

Der Fokus des Tests zur Multiplikation nach der Einführungsstunde lag auf dem konzeptuellen Verständnis zur Multiplikation (z.B. Wechsel zwischen Operationen Addition und Multiplikation;

einfache Multiplikationsaufgaben), welches eine bedeutende Vorläuferfähigkeit für Multiplikationsaufgaben darstellt.

Acht Tage nach der Einführungsstunde wurden die Schüler_innen erneut im Themenbereich Multiplikation getestet.

Die 12 im virtuellen Klassenzimmer vertretenen Schüler_innen wurden aus 579 Schüler_innen, die an beiden Testtagen teilgenommen haben, zufällig ausgewählt. Beachtet wurde dabei, dass die 12 Schüler_innen aus einer Klasse stammen. Die 12 Schüler_innen sind etwa gleichmäßig über die Leistungsbereiche verteilt.

Die Zuordnung der im virtuellen Klassenzimmer präsentierten Namen zu den jeweiligen Testheften erfolgte zufällig und es wurden hierfür fiktive Namen gewählt. Die verwendeten Namen wurden im Rahmen eines Pretests auf eine eindeutige Zuordnung zu einem Geschlecht und Migrationshintergrund vorgetestet.

Die Zweifensteransicht

Zunächst wird den Versuchspersonen eine kurze Erklärung der folgenden Zweifensteransicht, dem Kernstück des Mannheimer Inventars zur Testung der Urteilsakkuratheit präsentiert (Vgl. Abb.1)



Abbildung 1: Erklärung Zweifenster-Ansicht

Im nächsten Schritt erscheinen auf dem Bildschirm der Versuchspersonen zwei Fenster (die eigentliche Zweifensteransicht), eines links und eins rechts (Vgl. Abb. 4-Abb.10).

Die Versuchspersonen sehen durchweg beide Fenster und können daher während jeder Urteilsabgabe auf die Schülerinformation zurückgreifen.

Die Schülerinformationen

Im linken Fenster „Klassenzimmer“ befindet sich während der gesamten Urteilsabgabe die Ansicht einer virtuellen Schulklasse mit den 12 Schüler_innen. Die erste Ansicht enthält die Klassenübersicht mit den Namen: Murat, Emre, Farid, Max, Lukas, Jonas, Elif, Seda, Julia, Anne, Leonie und Anna (Vgl. Abb.2). Die Namen wurden im Rahmen einer Studie vorgetestet und sind eindeutig einer Nationalität sowie einem Geschlecht zu zuordnen.

Die Versuchspersonen erhalten die Information, dass es sich bei den Schüler_innen um Schüler_innen der zweiten Klasse handelt. Weiterhin werden die Versuchspersonen darauf hingewiesen, dass sie durch Anklicken auf den Namen weitere Informationen zum jeweiligen Schüler der jeweiligen Schülerin bekommen.

Durch Anklicken werden sodann ein Scan des Testheftes (22 Aufgaben), die erreichte Gesamtpunktzahl des Schülers/der Schülerin sowie die potentiell erreichbaren Punkte (Vgl. Abb.3) ersichtlich.

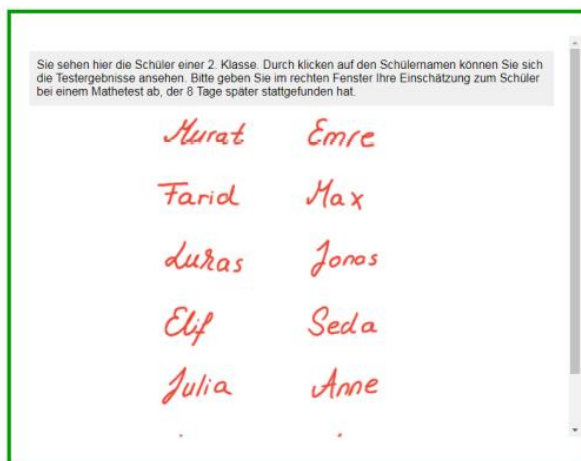


Abbildung 2: Klassenzimmer

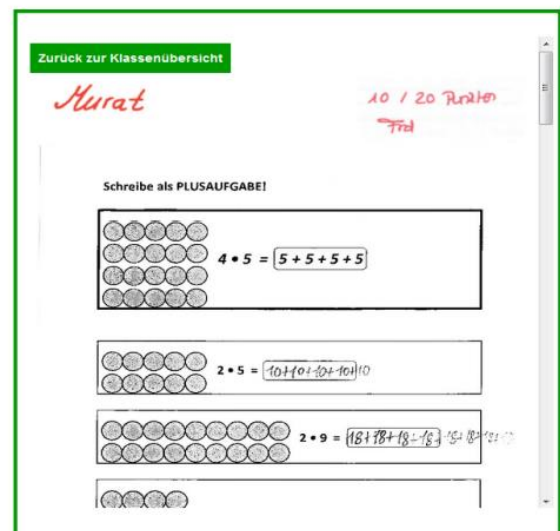


Abbildung 3: Beispiel Testheft

Die Schüler_innen variieren in ihrem Leistungsniveau. Es werden hierbei fünf Leistungsniveaus unterschieden (sehr gut, gut, durchschnittlich, weniger gut, schlecht) (Vgl. Tab.1). Diese Leistungsniveaus werden den Versuchspersonen nicht explizit genannt, sondern sind durch die Leistungen in den Testheften erschließbar. Die empirische Feststellung der Leistungsniveaus erfolgte auf der Grundlage von Quintilsanalysen über die Gesamtstichprobe hinweg.

Tabelle 1: Schülervariation

Name	Geschlecht	Migrations- hintergrund	Leistungsniveau ¹
Murat	♂		●●○○○
Emre	♂		●●●○○
Farid	♂		●●●●○
Max	♂		●●●●●
Lukas	♂		●●○○○
Jonas	♂		●●○○○
Elif	♀		●●●●●
Seda	♀		●○○○○
Julia	♀		●●●●○
Anne	♀		●○○○○
Leonie	♀		●●●●●
Anna	♀		●●●○○

¹ ●○○○○=niedriges Leistungsniveau ●●●●●=hohes Leistungsniveau

Die Urteile

Die abzugebenden Urteile (global/spezifisch/overall) befinden sich im rechten Fenster des Bildschirms der Versuchspersonen.

Sowohl die spezifischen Urteile als auch die globalen Urteile stellen Prognosen dar. Hier sollen die Versuchspersonen die Testleistung der Schüler_innen in einem Test zur Multiplikation 8 Tage später einschätzen.

Das Overall-Urteil bezieht sich auf eine Einschätzung der aktuellen Multiplikationsfähigkeiten der Schüler_innen auf Basis des vorliegenden Testheftes im linken Fenster.

Spezifisches Urteil:

Bei dem spezifischen Urteilsformat sehen die Versuchspersonen im rechten Fenster 11 Multiplikationsaufgaben von a) bis k) (Bsp: $2 \times 4 = \underline{\quad}$; Vgl. Abb. 4) mit der Aufforderung nun für jede Schülerin/ jeden Schüler anzugeben, welche dieser 11 Multiplikationsaufgaben die Schülerin/ der Schüler lösen kann. Hier können die Versuchspersonen für die Schüler_innen einzeln und für jede Aufgabe einzeln in einem Drop-Down Menü auswählen, ob der Schüler bzw. die Schülerin die richtige oder falsche Lösung angeben wird (dichotomes Antwortformat) (Vgl. Abb.5).

Abbildung 4: Spezifisches Urteil- Aufgabe

Sie sehen hier die Schüler einer 2. Klasse. Durch klicken auf den Schülernamen können Sie sich die Testergebnisse ansehen. Bitte geben Sie im rechten Fenster Ihre Einschätzung zum Schüler bei einem Mathetest ab, der 8 Tage später stattgefunden hat.

Aufgabe	Einschätzung
Aufgabe a: 2 x 8	-auswahl-
Aufgabe f: 4 x 8	-auswahl-
Aufgabe g: 7 x 8	-auswahl-
Aufgabe h: 2 x 22	-auswahl-
Aufgabe i: 3 x 33	-auswahl-
Aufgabe j: 4 x 25	-auswahl-
Aufgabe k: 3 x 66	-auswahl-

Anna

Aufgabe	Lösung Richtig oder Falsch?
Aufgabe a: 2 x 4	-auswahl-
Aufgabe b: 4 x 4	-auswahl-
Aufgabe c: 8 x 4	Richtig
Aufgabe d: 1 x 8	-auswahl-
Aufgabe e: 2 x 8	-auswahl-
Aufgabe f: 4 x 8	-auswahl-
Aufgabe g: 7 x 8	-auswahl-
Aufgabe h: 2 x 22	-auswahl-
Aufgabe i: 3 x 33	-auswahl-
Aufgabe j: 4 x 25	-auswahl-
Aufgabe k: 3 x 66	-auswahl-

Wieder

Abbildung 5: Spezifisches Urteil- Urteilsabgabe

Globales Urteil:

Das globale Urteilsformat teilt sich in drei verschiedene Urteile, die sich wiederum auf zwei verschiedene Aufgabentypen beziehen. Für die ersten beiden globalen Urteile (g1 und g2) sieht die Versuchsperson zwei identische Säulen, in deren Dach jeweils das Ergebnis (24 oder 40) einer Multiplikationsaufgabe zu sehen ist sowie die Aufforderung an die Schüler_innen eine Malaufgabe zu dem Ergebnis im Dach zu bilden. Die Aufgabe beinhaltet außerdem eine Beispiellösung für jede Aufgabe. Die Versuchsperson wird aufgefordert für jeden Schüler/jede Schülerin im offenen Antwortformat einzuschätzen, wie viele von den jeweils 7 möglichen richtigen Aufgaben für das Ergebnis 24 bzw. 40 von der Schülerin/ dem Schüler genannt werden (Vgl. Abb. 6 und 7).

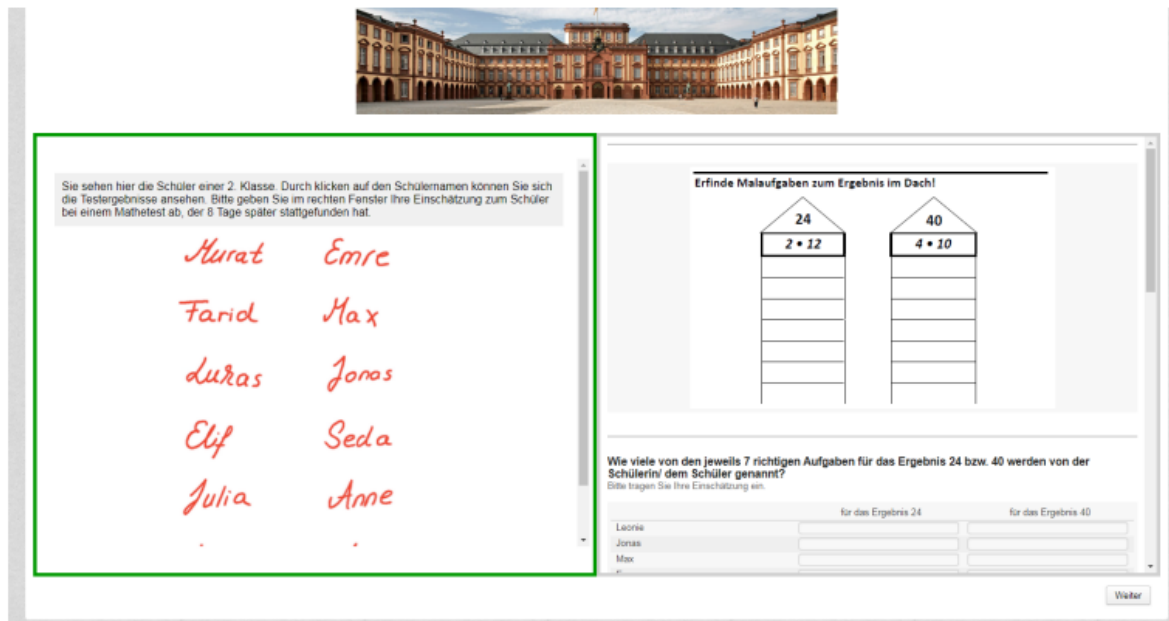


Abbildung 6: Globales Urteil (g1,g2)- Aufgabe

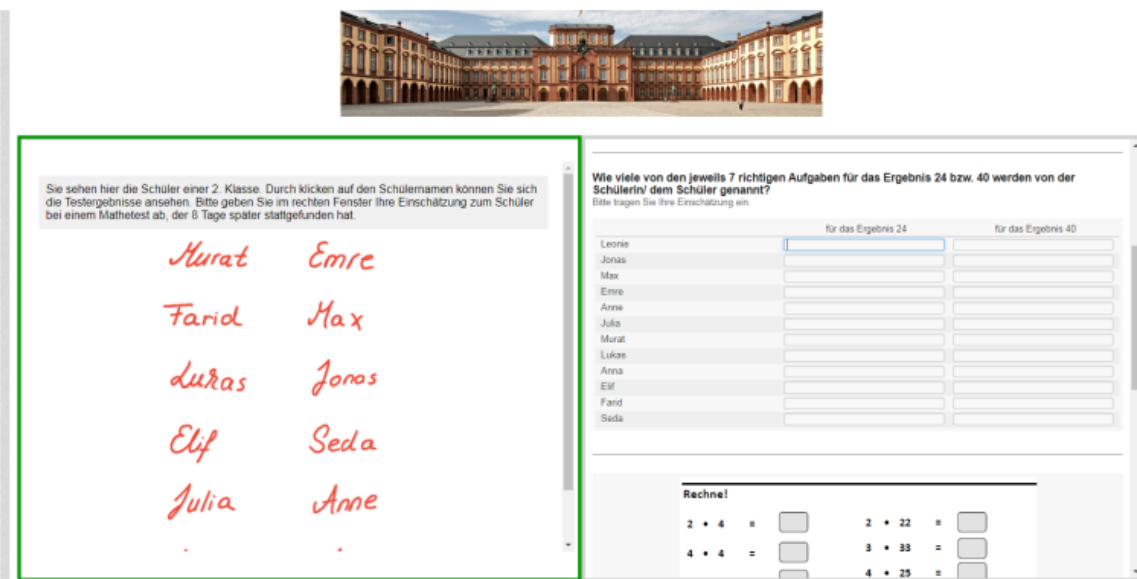


Abbildung 7: Globales Urteil (g1,g2)- Urteilsabgabe

Darüber hinaus werden die Versuchspersonen um ein weiteres globales Urteil gebeten (g3). Hier sehen die Versuchspersonen 11 Multiplikationsaufgaben mit der Aufforderung erneut in einem offenen Feld einzuschätzen, wie viele dieser 11 Multiplikationsaufgaben die Schülerin/ der Schüler lösen wird (Vgl. Abb. 8 und Abb. 9).

Sie sehen hier die Schüler einer 2. Klasse. Durch klicken auf den Schülernamen können Sie sich die Testergebnisse ansehen. Bitte geben Sie im rechten Fenster Ihre Einschätzung zum Schüler bei einem Mathetest ab, der 8 Tage später stattgefunden hat.

Murat Emre
Farid Max
Lukas Jonas
Elif Seda
Julia Anne

Rechne!

$2 \cdot 4 =$ $2 \cdot 22 =$
 $4 \cdot 4 =$ $3 \cdot 33 =$
 $8 \cdot 4 =$ $4 \cdot 25 =$
 $1 \cdot 8 =$ $3 \cdot 55 =$
 $2 \cdot 8 =$
 $4 \cdot 8 =$
 $7 \cdot 8 =$

Wie viele dieser 11 Multiplikationsaufgaben wird die Schülerin/ der Schüler lösen?
Bitte tragen Sie Ihre Einschätzung ein.

Leone

Weiter

Abbildung 8: Globales Urteil (g3)- Aufgabe

Sie sehen hier die Schüler einer 2. Klasse. Durch klicken auf den Schülernamen können Sie sich die Testergebnisse ansehen. Bitte geben Sie im rechten Fenster Ihre Einschätzung zum Schüler bei einem Mathetest ab, der 8 Tage später stattgefunden hat.

Murat Emre
Farid Max
Lukas Jonas
Elif Seda
Julia Anne

Rechne!

$7 \cdot 8 =$

Wie viele dieser 11 Multiplikationsaufgaben wird die Schülerin/ der Schüler lösen?
Bitte tragen Sie Ihre Einschätzung ein.

Leone

Jonas

Max

Emre

Anne

Julia

Murat

Lukas

Anna

Elif

Farid

Seda

Weiter

Abbildung 9: Globales Urteil (g3)- Urteilsabgabe

Overall-Urteil:

Bei diesem Urteilsformat handelt es sich um die Einschätzung des jetzigen Leistungsniveaus der Schüler_innen.

Auf der rechten Seite der Zweifensteransicht befindet sich in diesem Fall eine fünfstufige Skala mit den Skalenendpunkten „schlecht“ und „sehr gut“. Dies dient zur Einschätzung des Leistungsniveaus aller Schüler_innen mit der Instruktion: „Wie schätzen Sie insgesamt das Leistungsniveau des Schülers/ der Schülerin im Bereich Multiplikation ein?“ (Vgl. Abb. 10). Erneut haben die Versuchspersonen die Möglichkeit ihr Urteil nach Schüler_innen getrennt abzugeben.

Wie schätzen Sie insgesamt das Leistungsniveau des Schülers/ der Schülerin im Bereich Multiplikation ein?

	schlecht	weniger gut	durchschnittlich	gut	sehr gut
Leonie	☐	☐	☐	☐	☐
Jonas	☐	☐	☐	☐	☐
Max	☐	☐	☐	☐	☐
Emre	☐	☐	☐	☐	☐
Anne	☐	☐	☐	☐	☐
Julia	☐	☐	☐	☐	☐
Murat	☐	☐	☐	☐	☐
Lukas	☐	☐	☐	☐	☐
Anne	☐	☐	☐	☐	☐
Elif	☐	☐	☐	☐	☐
Farid	☐	☐	☐	☐	☐
Seda	☐	☐	☐	☐	☐

Viel Spaß

Abbildung 10: Overall Urteil

Kriteriumswerte

Das Overall-Urteil bezieht sich auf die Einschätzung der gegenwärtigen Schülerleistung. Hier dient als Kriteriumswert die tatsächliche Testleistung der Schüler_innen zu t1 (Testheft).

Das spezifische Urteil sowie das globale Urteil beziehen sich auf die Leistung der Schüler_innen 8 Tage später. Es handelt sich bei diesen Urteilen demnach also um Prognosen der Versuchspersonen. Hier gibt es zwei verschiedene Möglichkeiten, welche Form von Kriteriumswerte herangezogen werden. Zunächst können, wie zu erwarten, die tatsächlichen Schülerleistungen/ der tatsächliche Test-Score als Kriteriumswert herangezogen werden (Möglichkeit 1). Daneben gibt es aber auch die Möglichkeit (2), die auf Grundlage der Gesamtstichprobe ($n = 579$) zu erwartenden Testleistungen als unstandardisierte Erwartungswerte im Rahmen einer linearen Regression (mit der Gesamtestleistung zu t1 als

uV und der Leistung zu t_2 als aV) zu schätzen. Da es sich bei den vorliegenden Daten um Schülerleistungen von tatsächlichen Schülern handelt, welche in den zwischen den beiden Messzeitpunkten liegenden 8 Tagen weiter beschult wurden, ist diese Schätzung erwartbarer Werte nützlich, um die Prognose der Versuchspersonen unabhängig von für die Versuchspersonen nicht antizipierbare Ereignisse beurteilen zu können.

Generierte Daten

Mithilfe des virtuellen Klassenzimmers können die Antworten der Versuchspersonen gespeichert werden (mithilfe einer Erhebungssoftware, bspw. Unipark).

Darüber hinaus ist es mit dem virtuellen Klassenzimmer möglich Prozessdaten zu erheben (Dauer, Klickhäufigkeit auf Schüler_innen). Diese Prozessdaten werden im sogenannten „pupillog“ gesichert:

Für die Realisierung der virtuellen Klasse wurde eine Individualsoftware mit dem Namen „pupillog“ entwickelt und auf technischer Infrastruktur der Universität Mannheim betrieben. Diese Software lässt sich in jede webbasierte andere Erhebungssoftware integrieren. Für die bisherigen Erhebungen wurde Unipark als Erhebungssoftware verwendet.

Jede Navigation der Versuchsperson im virtuellen Klassenzimmer wird in Form eines Datenpunktes gespeichert. Zu Beginn sieht die Versuchsperson die Namen aller Schüler_innen. Der Versuchsperson ist es möglich zu navigieren, in dem sie den Namen eines Schülers bzw. einer Schülerin anklickt. Diese Aktion zeigt der Versuchsperson detaillierte Informationen zu dem gewählten Schüler bzw. der gewählten Schülerin an. Von dieser Detailansicht aus, kann die Versuchsperson nur zurück zur Übersicht mit allen Schülernamen navigieren. Ein direkter Wechsel zwischen Schüler_innen ist nicht möglich. Ein Datenpunkt besteht nun aus drei Werten: zum einen wird gespeichert, welcher Proband die Navigation ausgelöst hat. Nur so ist später eine Zuordnung zu den in einem anderen System (bspw. Unipark) erhobenen Daten zum gleichen Probanden möglich. Es wird weiterhin mit einer eindeutig zu einem Schüler bzw., einer Schülerin zugeordnete Nummer gespeichert, welchen Schüler, welche Schülerin, der Proband in der Detailansicht betrachtet. Das Navigieren zur Übersichtsseite mit allen Namen, wird ebenfalls mit einer festgelegten Nummer kodiert. Als dritter Wert wird für jede Navigation der genaue Zeitpunkt festgehalten. Betrachtet man nach Beendigung der Studie für einen Probanden alle seine Datenpunkte, ist mit den erhobenen Daten durch Differenzenbildung auch möglich zu ermitteln, wie lange ein Proband auf den einzelnen Seiten verweilt hat.

Zusammenfassend ist es möglich mithilfe des virtuellen Klassenzimmers die Zeit der Beschäftigung der Versuchspersonen mit den vorliegenden Informationen sowie das Klickverhalten probanden- und schülerspezifisch zu erfassen.

Publikationen, die auf MITU basieren:

- Bonefeld, M., Dickhäuser, O. & Karst, K. (2019). Do Preservice Teachers' Judgments and Judgment Accuracy depend on Students' Characteristics? The Effect of Gender and Immigration Background. *Social Psychology of Education*. Online First. doi:10.1007/s11218-019-09533-2.
- Karst, K., Bonefeld, M. & Siebert, J. (2018). Prozessdaten der Urteilsbildung und die Akkuratheit von Lehrkrafturteilen. Vortrag präsentiert auf dem 51. Kongress der Deutschen Gesellschaft für Psychologie (DGPS), Frankfurt am Main.
- Karst K., Bonefeld, M. & Dickhäuser, O. (2018). Judgment accuracy of pre-service teachers. Explained by pupil characteristics and attributes of the judge. Vortrag präsentiert auf der Conference: Cultural diversity, migration and education (CDME), Potsdam.
- Bonefeld, M., Karst, K. & Dickhäuser, O. (2018). Urteile über Schüler_innen mit Migrationshintergrund. Zur Bedeutung von Diagnose und Prognose. Vortrag präsentiert bei der 6. Tagung der Gesellschaft für empirische Bildungsforschung (GEBF), Basel.
- Karst, K., Bonefeld, M., Dickhäuser, O. (2017). Need for Cognition und die Akkuratheit von Urteilen angehender Lehrkräfte. Symposiumsbeitrag präsentiert auf der Tagung der Arbeitsgruppe für Empirische Pädagogische Forschung (AEPF), Tübingen.
- Bonefeld, M., Karst, K. & Dickhäuser, O. (2017). Leistungserwartungen von Lehrenden: Einfluss von Geschlecht und Migrationshintergrund der Lernenden. Symposiumsbeitrag präsentiert auf der Tagung der Arbeitsgruppe für Empirische Pädagogische Forschung (AEPF), Tübingen.
- Bonefeld, M., Karst, K. & Dickhäuser, O. (2017). Migrationshintergrund von Lernenden und Leistungserwartungen von Lehrkräften: Zur Bedeutung des Leistungsniveaus. Vortrag auf der gemeinsamen Tagung der Fachgruppen Entwicklungspsychologie und Pädagogische Psychologie (PAEPSY), Münster.

Literatur

- Lipowsky, F., Faust, G. & Kastens, C. (2013) (Hrsg.). Persönlichkeits- und Lernentwicklung an staatlichen und privaten Grundschulen. Ergebnisse der PERLE-Studie zu den ersten beiden Schuljahren. Münster: Waxmann.