

# FWM – Ein Test zur Erhebung von Wissen in Mechanik für Schülerinnen und Schüler der Mittelstufe

Tobias Ludwig

Pädagogische Hochschule Karlsruhe, Institut für Physik und Technische Bildung

Simon Zander

Albert-Einstein-Gymnasium, Kaarst

Heiko Krabbe

Ruhr-Universität Bochum, Didaktik der Physik

Burkhard Priemer

Humboldt-Universität zu Berlin, Didaktik der Physik

## Zusammenfassung

In Forschungsprojekten wird wiederholt Wissen in Mechanik – z. B. als Kontrollvariable – von Schülerinnen und Schülern der Mittelstufe erhoben. Dazu werden oft projektspezifische Tests entwickelt, die nicht in der Fachliteratur dokumentiert sind. Deshalb stellen wir einen Test zur ökonomischen Erhebung von Wissen aus Teilgebieten der Mechanik in der Mittelstufe vor, den FWM. Der aus 22 Items im Multiple-Choice Single-Select-Format bestehende Test kann in rund 20-25 Minuten bearbeitet werden. Wir stellen den Test sowie die Ergebnisse einer empirischen Erhebung mit 1532 Berliner Schülerinnen und Schülern aus der 8. und 9. Jahrgangsstufe vor. Die Ergebnisse zeigen, dass der Test in der Lage ist, Teilbereiche des heterogenen Konstrukts Wissen in Mechanik mit einer zufriedenstellenden Reliabilität zu erfassen. Eine Rasch-Analyse deutet zudem auf die Homogenität der Items hin. Mit diesem Beitrag möchten wir eine Skaldokumentation vorlegen und ein erprobtes Verfahren zur Erfassung von Wissen in Mechanik für zukünftige Forschungsvorhaben bieten.

Research projects often have the need to assess high school students' knowledge in mechanics, e. g. as a control variable. In Germany, most research projects develop own tests that usually are not published. Against this background, we present a test for an economic assessment of high school students' knowledge in certain mechanics topics, the FWM. The test consists of 22 multiple-choice single-select-items and can be administered in 20-25 minutes. We present the test and the results of a study with 1532 eighth- and nine-grade students in Berlin. The results show that the test can assess aspects of the heterogeneous construct "knowledge in mechanics" with a satisfactory reliability. A Rasch-analysis indicates the homogeneity of the items. With the paper, we want to provide a documentation of the instrument and offer researchers a field-tested tool to assess knowledge in mechanics.

*Keywords:* physics, content knowledge, mechanics, test

*Schlüsselwörter:* Physik, Fachwissen, Mechanik, Test, Mittelstufe

## 1. Themenfelder des Tests und Komplexität der Items

Wir haben ein Instrument – den FWM (Fachwissen Mechanik) – zur Erhebung von Teilaspekten des Wissens in Me-

chanik für die Mittelstufe (Klasse 7-10) nach einer Erstbegegnung mit Mechanik erstellt. Der Test wurde ursprünglich für Nordrhein-Westfalen entwickelt (Zander, Krabbe & Fischer, 2012) und für eine Studie in Nordrhein-Westfalen mit 30 Items und einer Testzeit von 45 Minuten verwendet (Zander, 2015, 2016). Danach wurde er überarbeitet und in der hier vorgestellten Form mit 22 Items in Berlin eingesetzt (Ludwig, 2017; Ludwig & Priemer, 2015). Der Test lässt sich mit dem Rasch-Modell skalieren und kann in rund 20-25 min bearbeitet werden. Mit diesem Beitrag möchten wir diesen verkürzten Test vorstellen, um ihn einer allgemeinen Nutzung zugänglich zu machen. Der Beitrag stellt eine Skalen-

---

Kontakt: tobias.ludwig@ph-karlsruhe.de.

Ludwig, T., Zander, S., Krabbe, H. & Priemer, B. (2021). *FWM - Ein Test zur Erhebung von Wissen in Mechanik für Schülerinnen und Schüler der Mittelstufe*. Karlsruhe: Pädagogische Hochschule Karlsruhe. Zugriff unter: <https://nbn-resolving.org/urn:nbn:de:bsz:751-opus4-2748>

dokumentation dar und verfolgt nicht das Ziel, Forschungsergebnisse vorzustellen. Abgedeckt sind vier Themenfelder der Mechanik: Kraft (im Folgenden abgekürzt durch K), Druck/Auftrieb (D), Arbeit / Energie / Leistung (E) und Bewegung (B), wobei Kraft den größten Umfang einnimmt (siehe Tabelle 1). Um zu den Themenfeldern Aufgaben mit unterschiedlichem Schwierigkeitsgrad zu generieren, wurde auf das Kompetenzmodell von Bernholt, Parchmann und Commons (2009) zurückgegriffen, das fünf Komplexitätsstufen definiert: Unreflektiertes Erfahrungswissen (U), Fakten (F), Prozessbeschreibungen (P), Lineare Kausalität (L) und Multivariate Interdependenz (M). Unter Prozessbeschreibungen haben wir qualitative Zusammenhänge erfasst, während quantitative Zusammenhänge der Stufe der linearen Kausalität zugeordnet wurden. Physikalische Konzepte fallen in die Stufe der multivariaten Interdependenz. Die vier Themenfelder sind mit mindestens drei Aufgaben auf mindestens zwei verschiedenen Komplexitätsstufen abgedeckt. Insgesamt sind alle Komplexitätsstufen in den Items vertreten, wobei im mittleren Bereich die meisten Items liegen (U = 1 Item, F = 4, P = 11, L = 4 und M = 2). Tabelle 2 beschreibt das dem Test zu Grunde liegende inhaltliche Konstrukt. Der Test ist auf Anfrage bei den Autoren erhältlich.

## 2. Empirische Evaluation des Tests

Um die psychometrische Qualität des Tests sowie die Verteilung der Itemschwierigkeiten über die Personenfähigkeiten zu bestimmen, wurde der Test in zwei Erhebungswellen 2014 und 2016 mit insgesamt 1532 Schülerinnen und Schülern von vier Berliner Gymnasien erprobt. Die Bearbeitungszeit der Schülerinnen und Schüler der 8. und 9. Jahrgangsstufe betrug etwa 20–25 min. 49 % der Probanden besuchten zum Zeitpunkt der Erhebung die 8. Jahrgangsstufe, die übrigen Probanden besuchten die 9. Jahrgangsstufe (Altersdurchschnitt  $14,1 \pm 0,8$  Jahre, 52 % Mädchen). Der Anteil fehlender Werte in den Itemantworten lag im Mittel bei 2,9 %. Auffällig ist hier, dass zwei Items (K09P01, 9,2 % und K07P01, 8,5 %) einen relativ hohen Anteil an fehlenden Werten aufwiesen. Beide Items adressieren den Umgang mit Kraftvektoren bei der grafischen Kräftezerlegung. Zur Auswertung des 22 Items umfassenden Tests wurden die Lösungen der Schülerinnen und Schüler dichotom kodiert. Im Mittel wurden 12,9 (58,7 %; Minimum drei Items, Maximum 21 Items) Items richtig beantwortet. Lösungshäufigkeiten je Item finden sich in Anhang A. Die interne Konsistenz nach Cronbachs wurde zu  $\alpha = .63$  bestimmt. Die Daten wurden zudem einer Rasch-Analyse unterzogen (siehe z.B. Bond & Fox, 2007; Boone, Staver & Yale, 2014; Neumann, 2014). Sämtliche Analysen wurden mit der Software R (R Core Team, 2014) durchgeführt. Neben Standardpaketen wurde das Paket TAM (Kiefer, Robitzsch & Wu, 2016) zur Schätzung des Rasch-Modells eingesetzt. Die Parameterschätzung erfolgte durch das Marginal-Maximum-Likelihood-Verfahren. Dabei

wurde der Mittelwert der Personenfähigkeitenverteilung auf null fixiert. Die beobachtete Varianz der Personenfähigkeiten beträgt 0,70. Die EAP-Reliabilität liegt bei ,63. Die mittlere Itemschwierigkeit liegt bei -0,44 Logits mit einer Varianz von 1,49 Logits. Die Itemschwierigkeiten liegen zwischen -2,97 und 2,74 Logits, die Standardfehler der Schätzer für die Itemschwierigkeiten liegen zwischen 0,05 und 0,11 Logits. Die Mean-Square (MNSQ) Fit-Statistiken der einzelnen Items liegen im Intervall von 0,88 bis 1,12 (Outfit) bzw. 0,93 und 1,08 (Infit) und entsprechen damit verbreiteten Cut-Off Kriterien ( $0,8 < \text{Infit} / \text{Outfit} < 1,2$ , Bond & Fox, 2007). Die t-standardisierten Werte der Itemparameter sind z. T. signifikant können aber ignoriert werden, da die MNSQ Fit-Statistiken innerhalb der Cut-Off Kriterien liegen (Boone et al., 2014, S. 166). Die Itemparameter und Fit-Statistiken sind in Anhang A dargestellt. In der Wright-Map sind die Personenfähigkeiten und Itemschwierigkeiten auf einer Skala aufgetragen (siehe Abbildung 1). Die Wright-Map zeigt eine gleichmäßige Abdeckung der Skala im Bereich von etwa -2,0 Logits bis +1,0 Logits mit einer kleinen Lücke zwischen -2,1 und -3,0 Logits, sowie zwischen 1 und 3 Logits.

## 3. Diskussion

Es konnte ein Rasch-Modell etabliert werden, welches anhand der Itemfit-Statistiken evaluiert wurde. Die Differenz der Mittelwerte der Verteilung der Itemschwierigkeiten bzw. Personenfähigkeiten lässt ferner den Schluss zu, dass der Test für die Stichprobe etwas zu leicht war. Dies zeigt sich auch an der Verteilung der Items in der Wright-Map. Es geht hervor, dass 14 von den 22 Items eine Schwierigkeit von  $< 0$  Logits aufweisen, d.h. dass 50 % der Probanden für diese Items eine Lösungswahrscheinlichkeit von mindestens 50 % aufweisen. Die Analyse der Wright-Map zeigt ferner, dass die Schwierigkeiten der Items recht gut über die Verteilung der Personenfähigkeiten gestaffelt sind. Insgesamt werden die Personenfähigkeiten zum größten Teil gut aufgelöst. Auffällig ist jedoch die Lücke im Bereich von 0,99 bis 2,74 Logits, die durch die Items K06P03 und K01P01 begrenzt wird (vgl. Abbildung 1). Dies deutet darauf hin, dass der Test in diesem Bereich durch weitere Items ergänzt werden könnte, um auch hier das Fachwissen in Mechanik differenziert aufzulösen. In diesem Bereich liegen jedoch nur 74 (4,8 %) der Probanden. Diese Lücke hat vermutlich auch einen Einfluss auf die mit ,63 schwache Reliabilität des Tests. „Zu große Lücken könnten die Reliabilität des Messinstruments beeinträchtigen, weil sich eine größere Menge an Personen (mit möglicherweise unterschiedlichen Fähigkeiten) nicht unterscheiden lässt“ (Neumann, 2014, S. 367). Auch die berichtete Personenvarianz scheint eher gering zu sein, was jedoch vergleichbar mit anderen fachdidaktischen Instrumenten der Kompetenzmessung ist (Kauertz, 2008; Neumann, 2014; Viering, 2012). Im Hinblick auf die Reliabilität muss ferner beachtet werden, dass mit den Items ein inhaltlich

Tabelle 1  
Identifizierte relevante Themenfelder der Mechanik für die Mittelstufe

<b>Kraft</b>		<b>Druck/Auftrieb</b>	
K01	Masse	D01	Druck, $P = \frac{F}{A}$
K02	Wirkung von Kraft	D02	Dichte, $\rho = \frac{m}{V}$
K03	Kraft als Vektor	D03	Schweredruck, $P = \rho \cdot g \cdot h$
K04	Kraftmessung (Hookesches Gesetz, $F = D \cdot s$ )	D04	Sinken, Schweben, Steigen, Schwimmen
K05	Ortsfaktor, $g = 9,81 \frac{N}{kg}$	D05	Luftdruck
K06	Gewichtskraft, $F_G = m \cdot g$	D06	Auftrieb, Auftriebskraft, $F_A = \rho \cdot g \cdot V$
K07	Kräftegleichgewicht	<b>Arbeit/Energie/Leistung</b>	
K08	Kräfteaddition	E01	Arbeit, $W = F \cdot s$
K09	Kräfteparallelogramm	E02	Lageenergie, $E_L = m \cdot g \cdot h$
K10	Kräftezerlegung	E03	Bewegungsenergie, $E_B = \frac{1}{2} \cdot m \cdot v^2$
K11	Wechselwirkungsgesetz	E04	Leistung, $P = \frac{W}{t}$
K12	Hebel, $F_1 \cdot l_1 = F_2 \cdot l_2$	E05	Energieumwandlung
K13	Flaschenzug, $F_Z = \frac{1}{n} F_G$	E06	Energieerhaltung
K14	Goldene Regel der Mechanik	<b>Bewegung</b>	
		B01	gleichförmige Bewegung
		B02	Geschwindigkeit, $v = \frac{s}{t}$
		B03	t-s-Diagramm zeichnen und interpretieren

Tabelle 2  
Beschreibung des vom Test erfassten inhaltlichen Konstrukts „Wissen in Mechanik“

<b>Themenfeld (Anzahl der Items, Komplexitätsstufen)</b>	<b>Schülerinnen und Schüler können...</b>
<b>Kraft</b> (13 Items; U, F, P und L)	...die Ursachen und Wirkungen von Kräften benennen, Bedingungen für ein Kräfte- bzw. Hebelgleichgewicht erkennen, die Richtung resultierender Kräfte bestimmen, die Komponentenerlegung von Kräften entlang vorgegebener Richtungen angeben, die Gewichtskraft auf einen Körper erkennen, die Gewichtskraft rechnerisch bestimmen bzw. daraus die Masse eines Körpers ableiten, das Verhältnis von Kräften und Wegstrecken bei einem Kraftwandler benennen
<b>Druck/Auftrieb</b> (3 Items; F und P)	...die Dichte verschiedener Körper vergleichen, Auftrieb in Flüssigkeiten und Luft aufgrund der Dichteverhältnisse begründen, Druckverhältnisse in Wasser bei verschiedenen Tiefen vergleichen
<b>Arbeit/Energie/Leistung</b> (3 Items; P, L und M)	...die Höhenlage eines Körpers mit der Lageenergie in Verbindung bringen, die beim Anheben eines Körpers verrichtete Arbeit vergleichen, die Formel für die Lage- und Bewegungsenergie kennen, Energieumwandlungen mit Hilfe der Energieerhaltung bilanzieren, den Begriff Leistung definieren
<b>Bewegung</b> (3 Items; F und M)	...eine Durchschnittsgeschwindigkeit berechnen und aus einem t-s-Diagramm ablesen, die damit in einer bestimmten Zeit zurückgelegte Strecke bestimmen

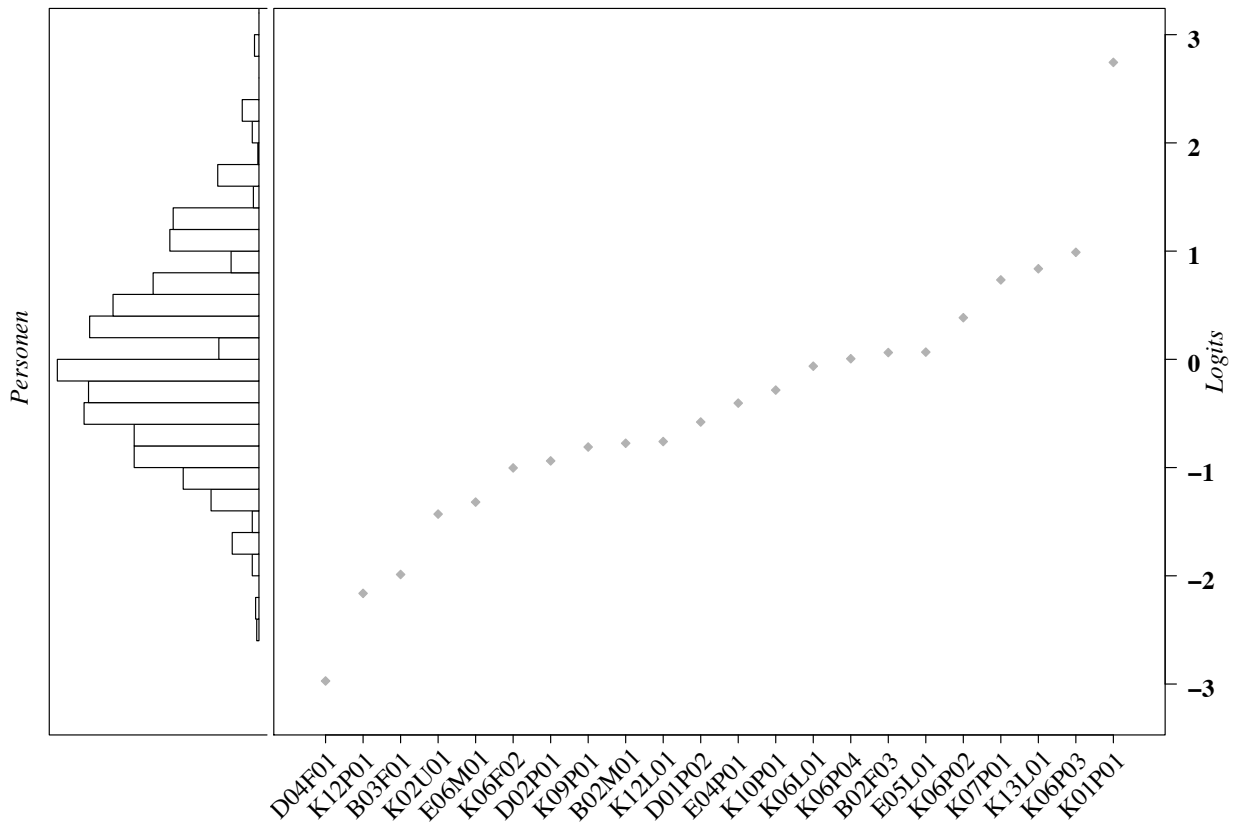


Abbildung 1. Wright Map (Gegenüberstellung von Personenfähigkeiten und Aufgabenschwierigkeiten) für 1532 Schülerinnen und Schüler unter Angabe der Häufigkeitsverteilung der Probanden. Die Items auf der Abszisse sind folgendermaßen kodiert: der erste Buchstabe steht für das Themengebiet, die zwei folgenden Zahlen stehen für Subthemen, der folgende Buchstabe beschreibt die Komplexitätsstufe und die letzten beiden Zahlen sind Itemnummierungen.

recht umfassendes Konstrukt erfasst wird (vgl. Tab. 1), dessen Teilthemen je nach Lehrkraft in verschiedenen Klassen unterschiedlich intensiv behandelt worden sein können bzw. für die Schülerinnen und Schüler unterschiedlich lange zurückliegen können. Mit dem vorgestellten Test (FWM) und dessen Testzeit von 20-25 Minuten stellen wir ein vergleichsweise ökonomisches Rasch-skalierbares Instrument zu Verfügung, um Wissen zu Teilgebieten der Mechanik ab der Mittelstufe im deutschen Bildungssystem reliabel zu erfassen. Bei der Verwendung des FWM sollte der Nutzer beachten, dass das abgefragte Konstrukt nur Teile des Inhaltsbereichs der Mechanik abdeckt. Ein Schwerpunkt bildet das Themenfeld „Kraft“. Eine vollständige Abdeckung des Curriculums kann durch den Test mit seiner begrenzten Anzahl an Aufgaben nicht erreicht werden. Neben einer Verwendung des Tests in seiner dargestellten Form kann dieser auch als Benchmark und Grundstock für projektspezifische Erweiterungen dienen. Wir empfehlen hinaus, die curriculare Validi-

tät und die Passung für andere Zielgruppen (z. B. Hauptschule, Realschulen oder Gesamtschulen) zu prüfen.

Veröffentlichung unter folgender Lizenz: Creative Commons - CC BY-NC-ND - Namensnennung - Nicht kommerziell - Keine Bearbeitungen 4.0 International <https://creativecommons.org/licenses/by-nc-nd/4.0/>



#### Literatur

- Bernholt, S., Parchmann, I. & Commons, M. L. (2009). Kompetenzmodellierung zwischen Forschung und Unterrichtspraxis. *Zeitschrift für Didaktik der Naturwissenschaften*, 15, 219–245.

- Bond, T. & Fox, C. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, N.J.: Lawrence Erlbaum Associates.
- Boone, W. J., Staver, J. R. & Yale, M. S. (2014). *Rasch Analysis in the Human Sciences*. New York, NY: Springer.
- Kauertz, A. (2008). *Schwierigkeitserzeugende Merkmale physikalischer Leistungstestaufgaben*. Berlin: Logos Verlag.
- Kiefer, T., Robitzsch, A. & Wu, M. (2016, 1. Februar). TAM: Test Analysis Modules (Version 1.16-0). Verfügbar 2. März 2016 unter <https://cran.r-project.org/web/packages/TAM/index.html>
- Ludwig, T. (2017). *Argumentieren beim Experimentieren - Die Bedeutung personaler und situationaler Faktoren* (Dissertation, Humboldt-Universität zu Berlin, Berlin). doi:10.18452/18408
- Ludwig, T. & Priemer, B. (2015). Argumentationen in Real- und Simulationsexperimenten: Ergebnisse einer experimentellen Laborstudie. In S. Bernholt (Hrsg.), *Heterogenität und Diversität - Vielfalt der Voraussetzungen im naturwissenschaftlichen Unterricht* (Bd. 35, S. 163–165). Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Bremen 2014. Kiel: IPN.
- Neumann, K. (2014). Rasch-Analyse naturwissenschaftsbezogener Leistungstests. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung*. Berlin Heidelberg: Springer-Verlag.
- R Core Team. (2014). R: A Language and Environment for Statistical Computing. Vienna, Austria. Verfügbar unter <http://www.R-project.org/>
- Viering, T. A. (2012). *Entwicklung physikalischer Kompetenz in der Sekundarstufe I. Validierung eines Kompetenzentwicklungsmodells für das Energiekonzept im Bereich Fachwissen*. Berlin: Logos-Verlag.
- Zander, S. (2015). Schülertest Mechanik Fach Physik. In H. E. Fischer (Hrsg.), *Instrumente fachdidaktischer Unterrichtsforschung Band 1*. Essen: Forschergruppe und Graduiertenkolleg nwu-essen. Verfügbar unter [http://duepublico.uni-duisburg-essen.de/servlets/DocumentServlet?Document=39989/Instrumente\\_fachdidakt\\_Unterrichtsforschung\\_Bd\\_1.pdf](http://duepublico.uni-duisburg-essen.de/servlets/DocumentServlet?Document=39989/Instrumente_fachdidakt_Unterrichtsforschung_Bd_1.pdf)
- Zander, S. (2016). *Lehrerfortbildung zu Basismodellen und Zusammenhänge zum Fachwissen*. Berlin: Logos Verlag.
- Zander, S., Krabbe, H. & Fischer, H. (2012). Entwicklung eines Fachwissenstests zur Mechanik für die Sekundarstufe I. In S. Bernholt (Hrsg.), *Konzepte fachdidaktischer Strukturierung für den Unterricht* (Bd. 32, S. 176–178). Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung in Oldenburg 2011. Berlin: LIT Verlag.

Anhang  
Kennwerte der Items

Tabelle A1

*Kennwerte der Items in der Reihenfolge des Testhefts (die Itemnamen haben folgende Bedeutung: der erste Buchstabe steht für das Themengebiet, die zwei folgenden Zahlen stehen für Subthemen, der folgende Buchstabe beschreibt die Komplexitätsstufe und die letzten beiden Zahlen sind Itemnummerierungen)*

Item	Richtige Lösung	Lösungshäufigkeit	$\xi$	$\xi_{SE}$	Outfit	Outfit <sub>t</sub>	Infit	Infit <sub>t</sub>	$r_{WLE}$
K12P01	4	0,88	-2,16	0,08	0,94	-0,79	0,98	-0,37	0,29
K02U01	2	0,79	-1,43	0,06	1,05	0,97	1,02	0,61	0,26
K06F02	4	0,71	-1,00	0,06	0,94	-1,88	0,97	-1,30	0,41
K09P01	2	0,68	-0,81	0,06	1,04	1,40	1,03	1,37	0,28
K10P01	1	0,56	-0,28	0,05	1,10	4,99	1,08	4,69	0,22
K12L01	2	0,67	-0,76	0,06	0,93	-2,68	0,95	-2,44	0,45
K06P04	4	0,50	0,01	0,05	1,02	1,17	1,02	1,10	0,35
K06L01	3	0,52	-0,06	0,06	1,06	3,39	1,05	3,37	0,28
K07P01	3	0,34	0,73	0,06	1,06	1,98	1,04	1,63	0,29
K13L01	1	0,32	0,84	0,06	1,02	0,54	1,01	0,32	0,35
K01P01	3	0,07	2,74	0,10	1,12	1,12	1,01	0,13	0,19
K06P03	4	0,29	0,99	0,06	0,91	-2,83	0,93	-2,62	0,49
K06P02	1	0,41	0,38	0,06	0,92	-3,96	0,93	-4,16	0,50
B03F01	4	0,86	-1,99	0,08	0,88	-1,77	0,97	-0,65	0,33
B02M01	3	0,67	-0,78	0,06	0,97	-1,05	0,98	-0,95	0,38
B02F03	2	0,49	0,06	0,05	0,94	-3,60	0,95	-3,68	0,47
D04F01	1	0,94	-2,97	0,11	0,91	-0,76	0,98	-0,17	0,23
D02P01	3	0,70	-0,94	0,06	0,91	-2,78	0,95	-2,10	0,44
D01P02	3	0,63	-0,58	0,06	1,00	0,11	1,00	0,19	0,36
E06M01	1	0,77	-1,32	0,06	1,03	0,63	1,02	0,49	0,29
E04P01	3	0,59	-0,40	0,06	1,09	3,98	1,07	3,96	0,24
E05L01	2	0,48	0,07	0,06	1,07	3,60	1,05	3,46	0,27

*Hinweise:  $\xi$ ...Itemschwierigkeit;  $\xi_{SE}$ ...Standardfehler der Itemschwierigkeit; Itemfitparameter MNSQ Outfit, Outfit<sub>t</sub>...t - Statistik für Outfit, Itemparameter MNSQ Infit, Infit<sub>t</sub>...t - Statistik für Infit;  $r_{WLE}$ ...punkt-biseriale Korrelation des Items mit der WLE-Personenfähigkeit (Trennschärfe)*