



MedAT

Kurzzusammenfassung 2013-2016

Martin Arendasy, Markus Sommer, Martina Feldhammer-Kahr



Inhaltsverzeichnis

Begründung der Entwicklung des MedAT-H und MedAT-Z 3

Anforderungen an die Qualität der beiden Aufnahmeverfahren und deren Erfüllung..... 3

 Objektivität und Fairness..... 3

 Messgenauigkeit..... 5

 Validität 6

Literaturverzeichnis..... 8

Begründung der Entwicklung des MedAT-H und MedAT-Z

Studiengangsspezifische Aufnahmeverfahren werden von StudienplatzwerberInnen nützlicher, fairer und moderner erlebt als alternative Methode zur Auswahl von StudienplatzwerberInnen (Hell & Schuler, 2005). Zudem kann ihre Nützlichkeit zur Vorhersage des späteren Studien- und Berufserfolgs in zahlreichen Studien und Metaanalysen empirisch belegt werden (z.B.: Hell, Trapmann, & Schuler, 2007; Kuncel, Credé, & Thomas, 2007; Kuncel & Hezlett, 2010; Kuncel, Hezlett, & Ones, 2001; Kuncel, Wee, Serafin, & Hezlett, 2010; Lievens, 2004; Reibnegger, Caluba, Ithaler, Manhal, Neges, Smolle, 2010, 2011; Schmidt-Atzert, 2005; Trost et al., 1998; Vetter & Sommer, 2012).

Den Ausgangspunkt der Entwicklung des MedAT-H und MedAT-Z stellte eine Evaluation des damals verwendeten Auswahlverfahrens dar (vgl. Spiel, Schober & Litzenberger, 2008). In dieser Studie zeigte sich, dass bei dem *früheren Aufnahmeverfahren Frauen und StudienplatzwerberInnen aus Österreich systematisch benachteiligt wurden, sodass sie auch im Falle gleicher Befähigung geringere Chancen hatten, aufgenommen zu werden als Männer und StudienplatzwerberInnen aus Deutschland*. Daher sollte ein neues Aufnahmeverfahren entwickelt werden, bei dem es empirisch nachweisbar zu keiner systematischen Benachteiligung unterschiedlicher Gruppen kommt.

Anforderungen an die Qualität der beiden Aufnahmeverfahren und deren Erfüllung

Die deutschen und amerikanischen Standards für pädagogisches und psychologisches Testen (AERA, APA & NCME, 1999; Häcker, Leutner, & Amelang, 1998) fordern bestimmte Mindeststandards für die Qualität von Aufnahmeverfahren hinsichtlich folgender Güte Merkmale:

1. Objektivität und Fairness
2. Messgenauigkeit
3. Validität

Objektivität und Fairness

Das Ergebnis eines Aufnahmeverfahrens ist mit Konsequenzen für den weiteren beruflichen Lebensweg der StudienplatzwerberInnen verbunden. Daher stellen die Objektivität und Fairness eines Aufnahmeverfahrens zentrale Gütekriterien dar (vgl. AERA, APA & NCME, 1999; Borsboom, 2006; Borsboom, Romeijn, & Wicherts, 2008; Drasgow, 1987; Häcker et al., 1998; Kunnan, 2000; Millsap, 2011; Mislevy et al., 2013; Wainer, 2002; Xi, 2010). Um die Objektivität und Fairness des MedAT-H und MedAT-Z zu gewährleisten wurden folgende Maßnahmen ergriffen:

1. Objektivität und Fairness durch einheitliche Rahmenbedingungen beim Aufnahmeverfahren
2. Objektivität und Fairness durch eine computergestützte Auswertung und Verrechnung
3. Objektivität und Fairness durch eine Sicherung eines gleichen Informationsstandes
4. Objektivität und Fairness durch jährliche, empirische Nachweise einer fehlenden Benachteiligung

An dieser Stelle soll nur kurz auf Maßnahmen und zentrale Befunde zu den letzten beiden Punkte eingegangen werden (für Details: Arendasy, Sommer, & Feldhammer-Kahr, 2016a).

Objektivität und Fairness durch Sicherung des gleichen Informationsstandes

Die deutschen und amerikanischen Standards für pädagogisches und psychologisches Testen sehen in den gleichen Möglichkeiten aller StudienplatzwerberInnen, sich mit Inhalten und Ablauf des Aufnahmeverfahrens vertraut zu machen, einen zentralen Aspekt der Fairness (AERA, APA & NCME, 1999; Häcker et al., 1998). Um dies zu gewährleisten, bieten die Medizinischen Universitäten in Kooperation mit der Abteilung für Psychologische Diagnostik und Methodik der Universität Graz kostenfreie Informationsbroschüren und Übungsaufgaben zu den einzelnen Aufgabengruppen des Aufnahmeverfahrens.

Objektivität und Fairness durch empirische Nachweise einer fehlenden Benachteiligung

Die deutschen und amerikanischen Standards für pädagogisches und psychologisches Testen fordern, dass bei einem Aufnahmeverfahren empirisch belegt werden muss, dass einzelne Gruppen an StudienplatzwerberInnen durch das Aufnahmeverfahren nicht systematisch bevorzugt oder benachteiligt werden (AERA, APA & NCME, 1999, Häcker et al., 1998).

Neben Studien zur Überprüfung einer fehlenden systematischen Benachteiligung einzelner Gruppen an StudienplatzwerberInnen im Vorfeld der Entwicklung von MedAT-H und MedAT-Z (z.B.: Arendasy & Sommer, 2011, 2012a, 2013c; Arendasy, Hornke, Sommer et al., 2009; Arendasy, Sommer, & Mayr, 2012) wurde jährlich im Anschluss an das Aufnahmeverfahren erneut untersucht, ob es durch die Aufgabengruppen der beiden Aufnahmeverfahren zu einer systematischen Benachteiligung einzelner Gruppen an StudienplatzwerberInnen kommt (vgl. Arendasy et al., 2013; Arendasy et al., 2014, 2015, 2016b, 2016c). Hierbei wurden folgende Schichtungsmerkmale berücksichtigt:

1. Geschlecht (Männer vs. Frauen)
2. Staatsbürgerschaft der StudienplatzwerberInnen (AT vs. EU-Staaten vs. Nicht-EU Staaten)

3. Muttersprache bzw. dominante Sprache der StudienplatzwerberInnen
4. Art der geplanten Finanzierung des Studiums (Stipendium und Nebenerwerb vs. Familie)
5. Bildungsstand der Eltern
6. Beruf der Eltern
7. Schultyp
8. Art der Vorbereitung auf das Aufnahmeverfahren
9. Vorerfahrung mit dem Aufnahmeverfahren (erstmaliger vs. wiederholter Testantritt)

In diesen Studien konnte belegt werden, dass einzelne Gruppen an StudienplatzwerberInnen durch das Aufnahmeverfahren nicht systematisch bevorzugt oder benachteiligt werden. *Dies bedeutet, dass zwei Personen mit gleicher Fähigkeit unabhängig von ihrer soziodemografischen Zugehörigkeit (vgl. Liste oben) die gleiche Wahrscheinlichkeit haben ein bestimmtes Testergebnis zu erreichen und aufgenommen zu werden (vgl. Arendasy et al., 2013; Arendasy et al., 2014, 2015, 2016b, 2016c).* Ergänzend wurde in begleitenden Forschungsprojekten der Medizinischen Universität Wien und der Abteilung für Psychologische Diagnostik und Methodik der Universität Graz auch belegt, dass prüfungsängstliche StudienplatzwerberInnen durch ihre Prüfungsangst bei MedAT-H und MedAT-Z nicht systematisch benachteiligt werden (vgl. Sommer & Arendasy, 2015; Sommer & Arendasy, 2016). Darüber hinaus konnte in einem weiteren begleitenden Kooperationsprojekt auch gezeigt werden, dass Unterschiede in der Art der Testvorbereitung nicht zu einer systematischen Bevorzugung von StudienplatzwerberInnen führen, die neben den kostenfreien Angeboten der Medizinischen Universitäten auch auf zusätzliche kommerzielle Testvorbereitungsangebote zurückgreifen (vgl. Arendasy, Sommer, Gutiérrez-Lobos, & Punter, 2016; Arendasy et al. 2016c).

Zusammenfassend belegen diese Befunde, dass es durch MedAT-H und MedAT-Z zu keiner systematischen Bevorzugung oder Benachteiligung einzelner soziodemografischer Gruppen an StudienplatzwerberInnen kommt, sodass zwei Personen mit gleicher Befähigung unabhängig von anderen Personenmerkmalen die gleiche Chance haben, einen bestimmten Testwert zu erzielen und aufgenommen zu werden.

Messgenauigkeit

Da die Entscheidung über die Aufnahme zum Studium mit weitreichenden Konsequenzen für die StudienplatzwerberInnen verbunden ist, fordern die deutschen und amerikanischen Standards für pädagogisches und psychologisches Testen (AERA, APA & NCME, 1999; Häcker et al., 1998) ein sehr hohes Maß an Messgenauigkeit für den Gesamtwert des Aufnahmeverfahrens. Dieses konnte für MedAT-H und MedAT-Z in den jährlich durchgeführten Studien belegt werden (vgl. Arendasy et al.,

2013; Arendasy et al., 2014, 2015, 2016b, 2016c). Die hier ermittelten Werte für die Messgenauigkeit der beiden Aufnahmeverfahren (MedAT-H: 0.93 bis 0.96 | MedAT-Z: 0.90 bis 0.98) lagen sogar über der in den Standards geforderten Messgenauigkeit.

Zusammenfassend kann daher festgehalten werden, dass es sich bei MedAT-H und MedAT-Z nicht nur um objektive und faire Aufnahmeverfahren handelt, sondern auch um messgenaue Aufnahmeverfahren.

Validität

Unter Validität bzw. Gültigkeit versteht man allgemein das Ausmaß, in dem ein Aufnahmeverfahren empirisch nachweisbar ein bestimmtes Merkmal erfasst, das auch zur Vorhersage eines interessierenden Verhaltens, wie beispielsweise dem Ausbildungs- und Berufserfolg relevant ist (für einen Überblick: Schmidt-Atzert & Amelang, 2012).

Als Belege für die Relevanz der einzelnen Aufgabengruppen liegen Metaanalysen und Studien mit einzelnen Aufgabengruppen des MedAT-H und MedAT-Z vor, in denen gezeigt werden konnte, dass die Aufgabengruppen der beiden Aufnahmeverfahren zur Vorhersage des Studienerfolgs beitragen (z.B.: Hell et al., 2007; Reibnegger et al., 2010, 2011; Trost et al., 1998; Vetter & Sommer, 2012). Als zentrales Ergebnis kann hier festgehalten werden, dass neben dem Vorwissen vor allem kognitive Fähigkeiten für die Prognose des Studienerfolgs relevant sind. Unter den kognitiven Fähigkeiten scheinen vor allem das Gedächtnis, das logisch-schlussfolgernde Denken, und die Raumvorstellung besonders relevant zu sein. *Während das Vorwissen in den ersten Semestern besonders stark zur Vorhersage des Studienerfolgs beitrug, nahm die Vorhersagekraft mit zunehmender Semesteranzahl ab. Im Gegensatz dazu nahm die Vorhersagekraft der Aufgabengruppen zur Messung der kognitiven Fähigkeit mit der Zeit zu und erwies sich vor allem für die Vorhersage der Leistung im United States Medical Licensing Exam (USMLE) als besonders relevant.*

Darüber hinaus liegen auch empirische Befunde vor, in denen belegt werden konnte, dass mit diesen Aufgaben tatsächlich die gewünschte Fähigkeit erfasst wird (z.B.: Arendasy & Sommer, 2011, 2012, 2013; Arendasy, Hornke, Sommer et al., 2009; Arendasy et al., 2012) und dass die einzelnen Aufgabengruppen mit den Testteilen Vorwissen (BMS), kognitive Fähigkeiten (KFF), Textverständnis (TV), manuelle Fähigkeiten (MF), und soziale Kompetenzen (SE) fünf voneinander klar unterscheidbare Determinanten des Studien- und Berufserfolgs erfassen (vgl. Arendasy et al., 2013; Arendasy et al., 2014, 2015, 2016c).

Zusammenfassend betrachtet, sprechen die vorliegenden Befunde somit auch für die Gültigkeit und Nützlichkeit der Aufgabengruppen des MedAT-H und MedAT-Z.

Literaturverzeichnis

American Educational Research Association, American Psychological Association and National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: Author.

Arendasy, M., Hornke, L. F., Sommer, M., Häusler, J., Wagner-Menghin, M., Gittler, G., et al. (2009). *Manual Intelligenz-Struktur-Batterie (INSBAT)*. Mödling: Schuhfried.

Arendasy, M., & Sommer, M. (2011). Automatisierte Itemgenerierung: Aktuelle Ansätze, Anwendungen und Forschungen. In L. F. Hornke, M. Amelang, & M. Kersting (Hrsg.), *Enzyklopädie für Psychologie: Methoden der Psychologischen Diagnostik* (S. 215-280). Göttingen: Hogrefe.

Arendasy, M., & Sommer, M. (2012). Using automatic item generation to meet the increasing item demands of high-stakes assessment. *Learning and Individual Differences, 22*, 112-117.

Arendasy, M., & Sommer, M. (2013). *Automatic item generation and first evidences on the dimensionality. Measurement fairness and construct representation of a picture completion task*. Unpublished research report. Graz: University of Graz.

Arendasy, M., Sommer, M., & Feldhammer, M. (2014). *MedAT-H & MedAT-Z 2014: Auswertungsdokumentation und psychometrische Evaluation*. Graz: Universität Graz.

Arendasy, M., Sommer, M., & Feldhammer, M. (2015). *MedAT-H & MedAT-Z 2015: Auswertungsdokumentation und psychometrische Evaluation*. Graz: Universität Graz.

Arendasy, M., Sommer, M., & Feldhammer, M. (2016a). *MedAT: Ausgangspunkt und erste Befunde*. Graz: Universität Graz.

Arendasy, M., Sommer, M., & Feldhammer, M. (2016b). *MedAT-H & MedAT-Z 2016: Auswertungsdokumentation und psychometrische Evaluation*. Graz: Universität Graz.

Arendasy, M., Sommer, M., & Feldhammer, M. (2016c). *MedAT-H Wien: Ergänzende Auswertungen zum sozio-ökonomischen Status, Schultyp, wiederholten Antritt und zur Muttersprache*. Graz: Universität Graz.

Arendasy, M., Sommer, M., Gutiérrez-Lobos, K., & Punter, J. F. (2016). Do individual differences in test preparation compromise the measurement fairness of admission tests? *Intelligence, 55*, 44-56.

Arendasy, M., Sommer, M., & Mayr, F. (2012). Using automatic item generation to simultaneously construct German and English versions of a word fluency test. *Cross Cultural Psychology, 43*, 464–479.

- Arendasy, M., Feldhammer, M., Reibnegger, G. & Sommer, M.(2013). *MedAT-H & MedAT-Z 2013: Auswertungsdokumentation und erste psychometrische Evaluation*. Graz: Universität Graz.
- Borsboom, D. (2006). When does measurement invariance matter? *Medical Care, 44*, 176-181.
- Borsboom, D., Romeijn, J.-W., & Wicherts, J. M. (2008). Measurement invariance versus selection invariance: Is fair selection possible? *Psychological Methods, 13*, 75-98.
- Drasgow, F. (1987). Study of measurement bias of two standardized psychological tests. *Journal of Applied Psychology, 72*, 19–29.
- Häcker, H., Leutner, D., & Amelang, M. (1998). *Standards für pädagogisches und psychologisches Testen*. Supplementum 1/2998 der Diagnostica und der Zeitschrift für Differentielle und Diagnostische Psychologie. Göttingen: Hogrefe.
- Hell, B., & Schuler, H. (2005). Verfahren der Studierendenauswahl aus Sicht der Bewerber. *Empirische Pädagogik, 19*, 361-376.
- Hell, B., Trapmann, S., & Schuler, H. (2007). Eine Metaanalyse der Validität von fachspezifischen Studierfähigkeitstests im deutschsprachigen Raum. *Empirische Pädagogik, 21*, 251-270.
- Kuncel, N. R., & Hezlett, S. A. (2010). Fact and fiction in cognitive ability testing for admissions and hiring decisions. *Current Directions in Psychological Science, 19*, 339-345.
- Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the Graduate Record Examinations: Implications for graduate student selection and performance. *Psychological Bulletin, 127*, 162–181.
- Kuncel, N. R., Credé, M., & Thomas, L. L. (2007). A Meta-Analysis of the predictive validity of the Graduate Management Admission Test (GMAT) and Undergraduate Grade Point Average (UGPA) for graduate student academic performance. *Academy of Management Learning & Education, 6*, 51-68.
- Kuncel, N. R., Wee, S., Serafin, L., & Hezlett, S. A. (2010). The validity of the Graduate Record Examination for master's and doctoral programs: A meta-analytic investigation. *Educational and Psychological Measurement, 70*, 340-352.
- Kunnan, A. J. (2000). Fairness and justice for all. In A. J. Kunnan (Ed.), *Fairness and validation in language assessment* (pp. 1–14). Cambridge, UK: Cambridge University Press.
- Lievens, F. (2004). Longitudinal study of the validity of different cognitive ability tests in a student admission context. *Applied H.R.M Research, 9*, 27–30.

Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York: Routledge.

Mislevy, R. J., Haertel, G., Cheng, B. H., Ructtinger, L., DeBarger, A., Murray, E., Rose, D., Gravel, J., Colker, A. M., Rutstein, D., & Vendlinski, T. (2013). A "conditional" sense of fairness in assessment. *Educational Research and Evaluation: An International Journal on Theory and Practice*, *19*, 121-140.

Reibnegger, G., Caluba, H.-C., Ithaler, D., Manhal, S., Neges, H., & Smolle, J. (2010). Progress of medical students after open admission or admission based on knowledge tests. *Medical Education*, *44*, 205-214.

Reibnegger, G., Caluba, H.-C., Ithaler, D., Manhal, S., Neges, H., & Smolle, J. (2011). Dropout rates in medical students at one school before and after the installation of admission tests in Austria. *Academic Medicine*, *86*, 1040-1048.

Schmidt-Atzert, L. (2005). *Vor- und Nachteile von Auswahlinstrumenten, Präsentation CHE Hochschulkurs*. http://www.hochschulkurs.de/downloads/Hochschulkurs/sf5_2005_schmidt-atz_stud.ppt

Schmidt-Atzert, L., & Amelang, M. (2012). *Psychologische Diagnostik und Intervention*. Berlin: Springer Verlag.

Sommer, M., & Arendasy, M. (2015). Further evidence for the deficit account of the test anxiety-test performance relationship from a high-stakes admission testing setting. *Intelligence*, *53*, 72-80.

Sommer, M., & Arendasy, M. (2016). Does trait test anxiety compromise the measurement fairness of high-stakes scholastic achievement tests? *Learning and Individual Differences*, *50*, 1-10.

Spiel, Ch., Schober, B., & Litzenberger, M. (2008). *Evaluation der Eignungstests für das Medizinstudium in Österreich*. Zusammenfassung und Empfehlungen. Universität Wien: Fakultät für Psychologie.

Trost, G., Blum, F., Fay, E., Klieme, E., Maichle, U., Meyer, M., & Nauels, H.-U. (1998). *Evaluation des Tests für medizinische Studiengänge (TMS): Synopsis der Ergebnisse*. Bonn: Institut für Test- und Begabungsforschung.

Vetter, M., & Sommer, M. (2012). *Ergebnisdokumentation zur prognostischen Validität des PMU Aufnahmeverfahrens*. Mödling: Schuhfried.

Wainer, H. (2002). On the automatic generation of items: Some whens, whys and hows. In S. H. Irvine, & P. C. Kyllonen (Eds.), *Item generation for test development* (pp. 287-316). New Jersey: Lawrence Erlbaum.

Xi, X. (2010). How do we go about investigating test fairness? *Language Testing*, *27*, 147-170.