

---

Zuwendungsempfänger:	Universität Konstanz
Vorhaben:	Genderfairness berufs- und studiendiagnostischer Tests
Förderkennzeichen:	01FP0930
Laufzeit:	01.10.2009 – 28.02.2014

---

## 1. Ergebnisse im Jahr 2012

**Das Vorhaben „Genderfairness berufs- und studieneignungsdiagnostischer Tests“ gliedert sich in die beiden Projektteile *Genderfairness von Verfahren zur Erfassung beruflicher Interessen (G-F BI)* und *Genderfairness fachspezifischer Studierfähigkeitstests (G-F SFT)*.**

**Im Projektteil *G-F BI* (Berufliche Interessen) wurden bisher drei Schwerpunkte abgearbeitet:**

- (1) Konzeptionelle Analyse der Genderfairness vorhandener Verfahren zur Erfassung beruflicher Interessen im deutschen Sprachraum: Dieser Projektteil wurde bereits im Jahr 2011 abgeschlossen. Die Ergebnisse sind im Jahresbericht 2011 dokumentiert.
- (2) Empirische Analyse der Genderfairness eines ausgewählten Verfahrens zur Erfassung beruflicher Interessen im Sinne des Hollandschen Hexagon-Modells: Die Ergebnisse verschiedener Methoden zur Untersuchung von Differential Item Functioning (DIF) zeigen übereinstimmend, dass sich auf allen sechs (RIASEC) Skalen des AIST-R Items befinden, die DIF enthalten und damit entweder Männer oder Frauen bevorzugen. Auf der Skala Realistic (praktisch-technische Interessen) wurden mit Abstand am meisten DIF-Items gefunden, insgesamt neun von zehn Items zeigten DIF. Bei den anderen Skalen gab es jeweils zwischen drei und vier DIF-Items. Aus diesen und weitergehenden Analysen lässt sich schließen, dass die Antworten der Untersuchungsteilnehmenden nicht nur durch ihre Ausprägung auf dem Konstrukt, welches die Items einer Skala eigentlich messen sollen (z.B. praktisch-technische Interessen) beeinflusst wird, sondern zusätzlich noch durch eine weitere Dimension, auf der sich Männer und Frauen unterscheiden.

Insgesamt zeigen unsere Ergebnisse, dass einige Items des AIST-R nicht messinvariant für Männer und Frauen sind. Besonders betroffen hiervon sind die Items der Dimension Realistic.

Mittelwertanalysen: Die Mittelwerte von Männern und Frauen auf den sechs AIST-R Skalen wurden verglichen. Es zeigten sich die üblichen Unterschiede: Männer berichteten ein größeres Interesse an praktisch-technischen und intellektuell-forschenden Tätigkeiten, während Frauen ein größeres Interesse an künstlerischen und sozialen Tätigkeiten angaben. Für den unternehmerischen und ordnend-verwaltenden Interessensbereich wurden keine nennenswerten Geschlechtsunterschiede festgestellt. Die Mittelwertunterschiede wurden basierend auf den Itemsets der experimentellen revidierten Fassung erneut untersucht. Für Realistic und Investigative verringerten sich die Unterschiede deutlich durch die Entfernung der Items mit substantiellen DIF.

Dennoch waren sie weiterhin vorhanden. Bei Artistic and Social dagegen blieben die Mittelwertunterschiede trotz Optimierung der Skalen durch Elimination von DIF-Items erhalten.

Analyse der differentiellen Validität und der differentiellen Prognose: Für den AIST-R konnten Geschlechtsunterschiede in der Zuordnungsgenauigkeit nachgewiesen werden. Männer wiesen in der Gesamtstichprobe einen signifikant höheren Passungsindex auf als Frauen. In der von uns entwickelten experimentellen DIF-reduzierten Testversion des AIST-R konnten diese Geschlechtsunterschiede nicht nachgewiesen werden. Insgesamt ist die DIF-reduzierte Testversion der originalen Testform tendenziell überlegen, da sie auch auf Itemebene weniger zwischen beiden Geschlechtern differenziert. Da der Aufwand für eine DIF-Reduzierung überschaubar ist, sollte diese Art der Testanalyse als Verfahren bei Interessentests standardmäßig angewendet werden, um geschlechtsspezifische Unterschiede zu minimieren.

Strukturanalysen: Genderübergreifende Gültigkeit des Hollandschen Strukturmodells. Es wurde untersucht, ob sich die Beseitigung von Items mit geschlechtsspezifischem DIF auf die psychometrische Struktur eines Standard RIASEC Interesses Inventars auswirkt. Die Ergebnisse deuten darauf hin, dass das Eliminieren von geschlechtsspezifischen DIF Items keinen nennenswerten Einfluss auf die psychometrische Struktur des Instruments hat.

Normierte oder nicht-normierte Summenwerte verwenden? Die Anwendung der Geschlechtnormen führt – wie nicht anders zu erwarten – zu einer deutlichen Veränderung der Skalenausprägungen gegenüber der Auswertung der Rohwerte. Betrachtet man die Ergebnisse zur Zuordnungsgenauigkeit, d.h. zur Frage nach der Personen-Umwelt-Passung in Form eines Kongruenzindex, ergeben sich signifikante Unterschiede zwischen den verschiedenen Auswertungsverfahren. Bei Verwendung von *Rohwerten* ergibt sich ein signifikant höherer Kongruenzindex als bei Verwendung von normierten Interessen. Werden die Analysen für Frauen und Männer separat durchgeführt, ergeben sich höhere Kongruenzmaße für Männer, wenn Gesamtnorm oder Rohwerte verwendet werden. Keine Unterschiede zeigen sich zwischen den Geschlechtern bei Anwendung der Geschlechtnorm.

Für eine Bewertung der Validität der Auswertungsverfahren ist allerdings die Frage entscheidender, ob sich die Auswertungsverfahren tatsächlich in ihrer Prognose der individuellen Zufriedenheit mit der Ausbildungs- und Studienwahl bzw. der individuell wahrgenommenen Passung der eigenen Interessen mit den Anforderungen des Ausbildungs- bzw. Studiengangs unterscheiden. Der Anteil der aufgeklärten Varianz ist für beide Kriterien bei Verwendung der Gesamtnorm bzw. der Rohwerte höher als bei Anwendung der Geschlechtnorm, allerdings erweisen sich diese Unterschiede nicht als signifikant. Somit kann aus diesen Ergebnissen keine eindeutige, abschließende Empfehlung für die Frage nach der optimalen Normierung von Interessentests gegeben werden. Zieht man aber die Ergebnisse der Kongruenzanalysen sowie die Vorteile, die mit der Verwendung von Rohwerten verbunden sind in Betracht, sprechen unsere Ergebnisse deutlich für die Verwendung von Testrohwerten. Der Konstruktionsaufwand ist wesentlich geringer (keine geschichteten Normen!) und Auswertung sowie Interpretation der Testergebnisse sind sowohl für die Testleitenden als auch für die Getesteten leichter und weniger fehleranfällig.

(3) Prüfung und Umsetzung von alternativen Testkonstruktionsmethoden zur Erstellung genderfairer Interessentests: Basierend auf den Ergebnissen der Pilotstudien wurde ein Interesseninventar bestehend aus deutschen Verben entwickelt. Das Interesseninventar wurde gemäß der im Antrag skizzierten Konstruktionsstrategien in drei Varianten entwickelt:

- VIT-Classic: Die erste Konstruktionsstrategie folgt den Regeln der klassischen Testtheorie. Zur Itemselektion wurden daher v.a. Itemschwierigkeiten und -trennschärfen berücksichtigt.
- VIT-Balanced: Die zweite Konstruktionsstrategie orientierte sich zwar im ersten Schritt auch an der klassischen Testtheorie, allerdings war es nun ein weiteres Ziel, möglichst ausbalancierte Skalen zu erstellen, indem Items mit großen Mittelwertdifferenzen zwischen Männern und Frauen entfernt wurden.
- VIT-IRT: Bei der dritten Konstruktionsstrategie wurde die Item Response Theorie herangezogen, die Schwellenparameter und die Wahrscheinlichkeiten der einzelnen Antwortkategorien wurden berücksichtigt. Zusätzlich wurden Items entfernt, die Differential Item Functioning zwischen Männern und Frauen aufwiesen

Somit liegen nun drei verschiedene Versionen des Interesseninventars vor, die in weiteren Erhebungen erprobt und anschließend revidiert wurden. Obwohl unterschiedliche Konstruktionsmethoden von uns eingesetzt wurden, zeigen die neu entwickelten Skalen doch eine gewisse Überlappung. Daher ist es auch nicht verwunderlich, dass auch die Ergebnisse hinsichtlich der Validität der Instrument-Varianten sich eher graduell unterscheiden. Die Konstruktvalidität aller drei Instrumente ist überzeugend: Die konvergenten Relationen zum AIST-R liegen in zufriedenstellender Höhe und die Skalen sind in allen Fällen unidimensional. Im Hinblick auf die Genderfairness erzielen die Instrumentvarianten VIT-IRT und VIT-Balanced die überzeugendsten Ergebnisse. Beide Instrumente regen die Exploration von nicht-stereotypen Ausbildungs-, Berufs- und Karriereoptionen an.

(4) Ergänzungsstudie: Konstruktion eines genderfairen Interessentests zur Ermittlung beruflicher Interessen von Schülern im deutschsprachigen Raum. Zur Konstruktion eines neuen genderfairen Interessentests wurden mehrere Projektschritte durchlaufen: (a) Erarbeitung eines ausdifferenzierten Interessenstrukturmodells, (b) Generierung von genderfairer, zielgruppenspezifischem Itemmaterial, (c) Erstellung einer ersten Testversionen, (d) Datenerhebung und Auswertung, (e) empirische Überprüfung des Interessenstrukturmodells sowie der statistischen Güte des Itemmaterials, (f) Erarbeitung von Studien- und Berufslisten zur Rückmeldung an Testteilnehmer, (g) Erstellung einer Onlineversion des Interessentests.

**Im Projektteil Genderfairness fachspezifischer Studierfähigkeitstest wird in drei Schwerpunkten gearbeitet.**

- (1) Schwerpunkt Metaanalyse: Der Projektschwerpunkt wurde im Jahre 2011 abgeschlossen. Die Ergebnisse sind im Bericht 2011 nachzulesen.
- (2) Fairnessuntersuchung prototypischer fachspezifischer Studierfähigkeitstests und Erarbeitung von Ansätzen zur Konstruktion genderfairer Verfahrensvarianten:

Konstruktvalidierung der fachspezifischen Studierfähigkeitstests: Für die Konstruktvalidierung der eingesetzten studienfeldbezogenen Beratungstestserien wurden Intelligenztestleistungen im logischen Schlussfolgern herangezogen. Dabei zeigt sich, dass die Studierfähigkeitstests in hohem Maße numerische Fähigkeiten abbilden. Zudem gab es einen klaren Zusammenhang zwischen den naturwissenschaftlichen Studierfähigkeitstests und verbalem Schlussfolgern.

Testinterne Analysen: Differential Item Functioning und Mittelwertanalysen: Die eingesetzten fachspezifischen Studierfähigkeitstests zeigen kein nennenswertes geschlechtsspezifisches Differential Item Function.

Analyse der differenziellen Validität und der differenziellen Prognose (auf verschiedenen Ebenen): Die Validitäten fallen für Frauen höher aus als für die Männer, sowohl auf der Ebene der Aufgabengruppen als auch auf Gesamtwert- und Konstruktebene. Der Unterschied ist jedoch nicht signifikant.

Die eingesetzten deutschen fachspezifischen Studierfähigkeitstests unterschätzen die Studienleistungen von Frauen auf Gesamttest-Ebene. Leistungsmotivation und Selbstdisziplin können die Unterschätzung von Frauen durch Studierfähigkeitstests teilweise erklären.

Es zeigt sich, dass die Aufgabengruppe „Naturwissenschaftliches Grundverständnis“ innerhalb der Naturwissenschaftlichen Stichprobe etwas weniger die Studienleistungen der Frauen unterschätzt. Dieses Ergebnis sollte für zukünftige genderfaire Testentwicklungen herangezogen werden.

- (3) Entwicklung eines „genderfairen“ kognitiven Fähigkeitstests:
  - Entwicklung neuer Leistungstest-Aufgabenformate: 198 Items wurden entwickelt (11 unterschiedliche Aufgabenformate mit jeweils 18 Items)
  - Überarbeitung der Items: Die Items wurden mehrfach aufgrund empirischer Analysen und theoretischen Überlegungen überarbeitet. Bei der Konstruktion und Überarbeitung aller Aufgabengruppen wurde darauf geachtet, dass sie keine Aufgaben enthalten, welche Frauen oder Männer benachteiligen (z.B. aufgrund des inhaltlichen Kontextes). Dies wurde mit DIF-Analysen (Differential Item Functioning) überprüft.
  - Zusammenstellung des genderfairen Tests: 6 der 11 Aufgabengruppen wurden in den Test übernommen, die sich systematisch voneinander unterscheiden und dementsprechend gruppieren lassen (verbal induktiv, figural induktiv, numerisch induktiv, verbal deduktiv, figural deduktiv, numerisch deduktiv) Die Endfassung des Tests besteht aus 96 Items und die Bearbeitung nimmt bei vollständiger Durchführung 75 Minuten in Anspruch.

## 2. Öffentlichkeitswirksames Material / Publikationsmaßnahmen 2012

### a) Internetauftritt und -präsentation

<http://www.uni-konstanz.de/genderfairness>

### b) Vorträge

Fischer, F.T., Schult, J. & Hell, B. (2012). Moderators of Sex-Specific Differential Prediction in College Admission Testing. Vortrag, 30. International Congress of Psychology, Cape Town, Südafrika, 22.-27.07.2012.

Päßler, K., Beinicke A. & Hell, B. (2012). Differential validity and differential prediction in interest inventories. Vortrag auf der 8. Conference of the International Test Commission, Amsterdam, Niederlande, 03.-05.07.2012.

Schult, J., Fischer, F.T. & Hell, B. (2012). How Intelligence Facets Relate to Admission Test Scores and Predict Academic Success. Vortrag, 30. International Congress of Psychology, Cape Town, Südafrika, 22.-27.07.2012.

Schult, J., Fischer, F.T. & Hell, B. (2012). Identifying Memory Strategies in Intelligence Testing Using Mixed Rasch Models. Vortrag auf dem 48. Kongress der Deutschen Gesellschaft für Psychologie, Bielefeld, 22.-27.09.2012.

### c) Informationsmaterialien (Flyer, Broschüren etc.)

--

### d) Medienarbeit (Fernsehen, Rundfunk, Tageszeitungen, *Fachzeitschriften*, Magazine, Pressemitteilungen etc.)

Fischer, F., Schult, J. & Hell, B. (in press). Sex Differences in Secondary School Success: Why Female Students Perform Better. *European Journal of Psychology of Education*. doi: [10.1007/s10212-012-0127-4](https://doi.org/10.1007/s10212-012-0127-4)

Fischer, F., Schult, J. & Hell, B. (in press). Sex-Specific Differential Prediction of College Admission Tests: A Meta-Analysis. *Journal of Educational Psychology*.

Päßler, K., Beinicke, A. & Hell, B. (in press). Gender related Differential Validity and Differential Prediction in Interest Inventories. *Journal of Career Assessment*.

Schult, J., Hell, B., Päßler, K. & Schuler, H. (in press). Sex-Specific Differential Prediction of Academic Achievement by German Ability Tests. *International Journal of Selection and Assessment*.

Päßler, K. & Hell, B. (2012). Do Interests and Cognitive Abilities Help Explain College Major Choice Equally Well for Women and Men? *Journal of Career Assessment*, 20, 479-496. doi: [10.1177/1069072712450009](https://doi.org/10.1177/1069072712450009)

Wetzel, E. & Hell, B. (2012). Sind deutschsprachige Interessentests gender fair? - Eine qualitative Analyse. *Zeitschrift für Arbeits- und Organisationspsychologie*, 56, 37-47. doi: [10.1026/0932-4089/a000065](https://doi.org/10.1026/0932-4089/a000065)

Wetzel, E., Hell, B. & Päßler, K. (2012). Comparison of different test construction

strategies in the development of a gender fair interest inventory using verbs. *Journal of Career Assessment*, 20 (1), 86 - 102. [doi: 10.1177/1069072711417166](https://doi.org/10.1177/1069072711417166)

e) Werbung (Plakate, Anzeigenkampagne etc.)

--

f) Marketingartikel (Blöcke, Kugelschreiber, Lesezeichen, Schlüsselanhänger etc.)

--

g) Sonstiges

--