Kubinger, Klaus (2009). Zusammenfassung zur Diplomarbeit Ingrid Preusche. (Link erloschen. Download am 1. Mai 2011 unter  http://psychologie.univie.ac.at/eppd/studium/psychologische-diagnostik/diplomarbeiten/ diplomarbeiten-seit-1998/kubinger/preusche/)

***Fairness von Intelligenztests bei Kindern und Jugendlichen***

Intelligenztests sind seit Jahrzehnten unabdingbare Instrumente u.a. im Zusammenhang mit allen Fragestellungen zur psychologischen Schul-, Berufs- und Laufbahnberatung. Gerade bei Kindern und Jugendlichen erweist sich in diesem Zusammenhang, nach allgemeiner übereinstimmender Meinung, eine Einzeltestung einer Gruppentestung vorzuziehen. Erst durch die besondere Beobachtungsmöglichkeit bzw. durch eine Diagnostik der Interaktion der Tp mit dem Tl können relevante Informationen erfasst werden. Allerdings birgt die intensive Interaktion von Tp und Tl die Gefahr systematischer Testleitereffekte. Und diese werden in der Fachliteratur nicht genauer erforscht, wohl aber kritisch reklamiert.

Frau Preusche widmete sich daher in ihrer Dissertation diesem Thema. Ihr Augenmerk galt den beiden wohl zwischenzeitlich im deutschsprachigen Raum alle Fallbehandlungen nahezu komplett abdeckenden Verfahren AID 2 (Adaptives Intelligenz Diagnostikum - Version 2.1; Kubinger & Wurst, 2000) und HAWIK III (Hamburg-Wechsler Intelligenztest für Kinder III; Tewes, Rossmann & Schallberger, 1999). An Faktoren, die alle einen Einfluss auf das Testergebnis eines Kindes haben könnten, untersuchte sie neben diversen denkbaren Testleitereffekten das in der Praxis besonders relevante Problem der Testwiederholung.

Der besonders komplizierte Versuchsplan sah vor allem vor, den Effekt der Testwiderholung unspezifisch (von HAWIK III auf AID 2 - in bezug auf inhaltlich ähnliche Untertests - einerseits und AID 2 auf HAWIK III andereseits) zu untersuchen, spezifisch (von AID 2 auf AID 2 mit individuell angepasster Parallelform) und "meta"-spezifisch (von AID 2 auf AID 2 in der Standardform). Sodann war vorgesehen, Unterschiede zwischen einzelnen Tl zu quantifizieren, Unterschiede zwischen dem Geschlecht der Tl und vor allem geschlechtsspezifische Interaktionseffekte in bezug auf Tl und Tp - schließlich gibt es seit Hawel (1979) den empirischen, seither nie wieder untersuchten Befund, dass Mädchen im HAWIK bei männlichen Tl besser als bei weiblichen Tl abschneiden.

<>

<>Im Vorfeld wurde inhaltlich festgelegt, ein wie großer signifikanter Unterschied als relevant zu betrachten ist: Differenzen zwischen 4 und 5 T-Werten wurden als "grenzwertig bedeutsam" definiert, Differenzen über 5 T-Werte als bedeutend, weil konsequenzreich.

Bei der Analyse der Daten zeigte sich, dass ein zusätzlicher Faktor die Ergebnisse der Kinder beeinflusst: Die Schulzugehörigkeit (retrospektiv am ehesten operationalisiert als schichtspezifisch unterschiedlich geprägte politische Bezirke). So mussten die Schulen in zwei Kategorien gruppiert werden, die eher leistungsstarke vs. eher leistungsschwache Kinder repräsentierten. Dementsprechend wurden die Auswertungen zumeist pro Schulgruppe vorgenommen. Generell zeigten sich für beide Gruppen oft divergierende Ergebnisse, sodass die untersuchten Einflussfaktoren (leistungs-) niveauspezifisch zu wirken scheinen.

<>

<>Bei Kodieraufgaben (Zahlen-Symbol-Test bzw. Kodieren und Assoziieren) zeigen sich deutliche Lerneffekte. Und beispielsweise beim Untertest "Soziale und Sachliche Folgerichtigkeit" zeigt sich, dass die vorhandene Parallelform im AID 2 einen Übungs-/Lerneffekt bei wiederholter Testung weitgehend verhindern kann. Bei beiden Testbatterien konnten/mussten Testleitereffekte nachgewiesen werden, allerdings lediglich bei eher leistungsstarken Kindern. Am auffälligsten sind Testleitereffekte in den AID 2-Untertests Synonyme Finden und Kodieren und Assoziieren sowie im IQ des HAWIK III (bis zu 13 Wertpunkten in der Summe aller Wertpunkte). Genauer, aber nur deskriptiv wurde der Testleitereffekt auf Itembasis für den Wortschatztest des HAWIK-III untersucht: Aufgrund von unterschiedlichen Bewertungen, die u.a. auf Uneindeutigkeiten und Widersprüchlichkeiten in der Bewertung laut Manual zurückzuführen sein dürften, zeigten sich innerhalb der ersten acht Items deutliche Testleitereffekte, die sich allerdings nicht in einer systematischen Unter- oder Überbewertung seitens ein und desselben Tl ausdrücken.

Die Ergebnisse hinsichtlich des Einflusses des Geschlechts der Tl sind nicht einheitlich: Immerhin zeigt sich bei den eher leistungsstarken Kindern, dass weibliche Tl bei ihren Tpn bessere Ergebnisse in den Untertests Rechnerisches Denken bzw. Angewandtes Rechnen erzielen. Bei der Gruppe der eher leistungsschwachen Kindern scheint eine Wechselwirkung im AID 2 beim minimalen T-Wert im Ausmaß von 6 T-Werten zu bestehen, und zwar zugunsten der Kombination "männlicher Tl testet weibliches Kind".