



Gebräuchliche Normen: z-Werte (-3 - +3), T-Werte (20 - 80), IQ des HAWIE (55 - 145), SW des IST (70 - 130), C-Werte (0 - 10), Stanine (1 - 9)

Jede lineare Transformation statistisch gerechtfertigt

z [Eichmaßstab] kann die relative Position einer Vp hinsichtlich der entsprechenden Referenzpopulation adäquat beschreiben

Flynn-Effekt: Durchschnittwert des IQs steigt innerhalb von ca. 20 Jahren bis zu 25 IQ-Punkte an

Angemessenheit der Normwerte ist spätestens alle 8 Jahre zu prüfen (DIN 33430)

Aktualität der Eichtabellen

Eichung im Sinn von Repräsentativerhebung

Stichprobe gibt ein verkleinertes Abbild der Population wieder

Stichprobengröße ist weitgehend irrelevant, besonders große Stichproben sind keine Garantie für Repräsentativität

Anfallstichproben sind sehr wahrscheinlich verzerrt (z.B. Freiwilligenstichproben)

Klumpenstichproben

Vollerhebung einzelner „Klumpen“ von Datenträgern

Klumpen zufällig gewählt

Kriteriumsorientierte Diagnostik

Absolut festgelegter oder individuumsbezogener, relativ gewählter Ziel-Testwert

Heranziehen anderer Vergleichsmaßstäbe zur Interpretation eines Testwerts als die Verteilung in der Lernerfolg, Therapieerfolg

IRT

Regelmäßig erweisen sich Tests als nicht verrechnungsfair

Im HAWIK-R stellte sich heraus, dass einige Items in Abhängigkeit vom Geschlecht unterschiedliche (relative) Schwierigkeiten besitzen

Tests, für die das Rasch-Modell nicht gilt, sind nicht verrechnungsfair

Kann prüfen, ob die gegebenen Verrechnungsvorschriften eines Tests zu Testwerten führen, die verhaltensadäquate Relationen wiedergeben

Rasch-Modell muss gelten (Wahrscheinlichkeit, dass Vp Item i löst ist abhängig von Personenparameter, wahre Fähigkeit und Itemparameter, wahre Schwierigkeit)

Rasch-Modell ist prüfbar (im Gegensatz zur Faktorenanalyse), weil stichprobenunabhängig

NEBENGÜTEKRITERIEN

Wird dieselbe Eigenschaftsdimension gemessen?

Stimmt für beide Fälle der Mittelwert der Testwerte

Stimmt für beide Fälle die Verteilung der Testwerte (Standardabweichung) überein?

Übertragung in eine andere Kultur

Skalierung

Resultierende Testwerte bilden die empirischen Verhaltensrelationen adäquat ab

Verrechnung der Testleistungen und -reaktionen zu einem numerischen Testwert

KTT

Schwierigkeiten der Items verteilen sich gleichmäßig zwischen 0,05 und 0,95

Trennschärfeindizes (Korrelationen des Testwert pro Item mit dem Testwert aus allen übrigen Items) nehmen sehr hohe Werte an

Kritik: Methoden der KTT sind stichprobenabhängig

Lösungshäufigkeiten p ändert ihren Wert drastisch in Abhängigkeit von der Zusammensetzung der Stichprobe

Korrelation der Testwerte kann für zwei Teilstichproben gänzlich andere Werte annehmen als für die Gesamtstichprobe

Durch geeignete Stichprobenauswahl ist praktisch jeder gewünschte Wert erreichbar (Manipulierbarkeit)

Ökonomie

Wirtschaftlichkeit, Aufwandsminimierung

Gemessen am diagnostischen Infogewinn beansprucht der Test relativ wenig Ressourcen

Kostenfaktoren

Anschaffung eines Tests

Verbrauch von Protokollbogen bzw. der Gebühr für die Nutzung computerisierter Tests

Personalkosten des TL

Anschaffungs- und Betriebs-, Verschleißkosten von Computer

Eichen im Sinn von Relativieren

Die für die herangezogene Stichprobe ist repräsentativ

Die Normen sind gültig, d.h. nicht veraltet

Die Population, für die die Normen gelten, ist definiert

Allgemeines

Eichung