

# 1 KURZBESCHREIBUNG DES VERFAHRENS

## Autoren

M. Arendasy, M. Sommer & G. Gittler

## Anwendung

Erfassung des Intelligenzniveaus und der Intelligenzstruktur; einsetzbar ab 14 Jahren.

Hauptanwendungsbereiche: Eignungsdiagnostik, Personalpsychologie, Pädagogische Psychologie, Schul- und Berufsberatung sowie Flugpsychologie

## Theoretischer Hintergrund

Die Intelligenz-Struktur-Batterie (INSBAT-2) stellt eine Weiterentwicklung der INSBAT dar. Im Sinne einer entscheidungsorientierten psychologischen Diagnostik ist INSBAT-2 daher ebenfalls modular aufgebaut. Dies bedeutet, dass nur jene Subtests vorgegeben werden müssen, die zur Beantwortung der Fragestellung erforderlich sind.

Die INSBAT-2 basiert auf dem Cattell-Horn-Carroll Modell (CHC-Modell: McGrew, 1997, 2009; Schneider & McGrew, 2018). Dieses Modell nimmt an, dass sich die Zusammenhänge zwischen den einzelnen Subtests durch eine Reihe hierarchisch angeordneter Intelligenzfaktoren erklären lassen. Von den mittlerweile 17 theoretisch postulierten Sekundärfaktoren des CHC-Modells werden mit Hilfe der INSBAT-2 insgesamt 5 Sekundärfaktoren durch jeweils zwei bis drei Subtests erfasst. Hierbei handelt es sich um die Fluide Intelligenz ( $G_f$ ), die Kristalline Intelligenz ( $G_c$ ), das Quantitative Denken ( $G_q$ ), die Visuell-räumliche Verarbeitung ( $G_v$ ), und das Langzeitgedächtnis ( $G_{lr}$ ). Die ausgewählten fünf Sekundärfaktoren konnten nicht nur in einer Vielzahl faktorenanalytischer Studien wiederholt repliziert werden, sondern erwiesen sich auch in aktuellen Metanalysen als prädiktiv valide für die Vorhersage des Ausbildungs- und Berufserfolgs.

Insgesamt stehen somit 12 Subtests zur Verfügung. Die Aufgaben der einzelnen Subtests wurden mit Hilfe der Automatisierten Itemgenerierung (AIG: Arendasy & Sommer, 2011; Irvine & Kyllonen, 2002) unter Verwendung aktueller Forschungsergebnisse aus den Kognitions- und Neurowissenschaften, der Differentiellen Psychologie, und der Angewandten Psychometrie konstruiert. Für die Subtests ergaben sich in verschiedenen Studien Belege für die Passung des 1PL Rasch Modells. Darüber hinaus sprechen die Befunde auch für die Geschlechts-, Bildungs-, und Altersfairness der Subtests, sowie für deren Konstrukt- und Kriteriumsvalidität. Hieraus ergeben sich für Anwender des Verfahrens folgende Vorteile: (1) Verrechnungsfairness, (2) Interpretationseindeutigkeit der Testwerte, und (3) prognostische Relevanz im Bereich der Berufs- und Bildungsberatung bzw. im Bereich der Personalauswahl und Zulassung zu unterschiedlichen Studien- und Ausbildungsgängen.

## Durchführung

Die INSBAT-2 wurde als modulare Intelligenztestbatterie entwickelt, sodass im Prinzip nur jene Subtests vorgegeben werden müssen, die zur Beantwortung der diagnostischen Fragestellung relevant sind. Anwender können nicht nur die für ihre diagnostischen Fragestellungen relevanten Subtests auswählen, sondern auch die Reihenfolge der Subtests, sowie die Endbedingungen der adaptiven Tests an die Fragestellung anpassen.

Für alle Subtests liegt eine standardisierte Instruktion mit Übungsbeispielen vor, die auf Prinzipien des programmierten Unterrichts und „mastery learning“ basieren. Die Antworteingabe erfolgt je nach Subtest im Multiple-Choice-Format oder in automatisierter freier Beantwortung. Bei allen Subtests werden die Aufgaben als adaptiver Test (CAT) mit einem anhand soziodemografischer Daten optimierten Testeinstieg vorgegeben, wodurch die gewonnene Information ohne Überforderungen bzw. Unterforderung der Testpersonen durch einzelne Aufgaben maximiert wird.

## Testformen

Es stehen zwei Testformen zur Verfügung, die sich hinsichtlich ihrer Möglichkeiten zur Subtestauswahl und Darbietung unterscheiden. Die Standardform S1 besteht aus zwölf Subtests und eignet sich für eine umfassende Leistungserfassung. In dieser Testform stehen alle zwölf Subtests zur Verfügung. Die Auswahl der Subtests, sowie des Vorgabemodus (Adaptiv vs. Linear) kann hier wie gewohnt gewählt, und an die diagnostische Fragestellung angepasst werden. Aus Gründen der Testsicherheit ist diese Testform (S1) nur für eine Administration unter Aufsicht vorgesehen. Die Screeningform S2 besteht aus drei Subtests (Figural-induktives Denken, Verbale Flüssigkeit und Mathematische Flexibilität) und kann für ein Testen ohne Aufsicht genutzt werden. Beide Testformen können adaptiv oder als lineare Testform mit randomisierter Itemauswahl vorgegeben werden.

## Auswertung

Für jeden der ausgewählten Subtests wird der Fähigkeitsparameter nach dem Rasch-Modell ausgegeben. Während die Anzahl der richtig bearbeiteten Items lediglich ein Maß für die Performanz einer Person darstellt, erlaubt der Fähigkeitsparameter die Schätzung der zugrunde liegenden, latenten Fähigkeitsdimension. Dies stellt einen wesentlichen und zugleich auch diagnostisch notwendigen Inferenzschritt dar. Dieser Inferenzschritt ist jedoch an die Passung des Rasch-Modells gebunden (vgl. dazu van der Linden & Hambleton, 1997), die für die Subtests der INSBAT-2 belegt werden konnte.

Sofern zu einem Sekundärfaktor mehr als ein Subtest ausgewählt wurde, wird ein Faktorwert für den entsprechenden Sekundärfaktor berechnet. Der Faktorwert für den Generalfaktor der Allgemeinen Intelligenz (G) wird berechnet, wenn zumindest vier Subtests aus unterschiedlichen Sekundärfaktoren vorgegeben wurden. Neben der Ausgabe der Fähigkeitsparameter und der Faktorwerte wird zudem automatisch ein Normvergleich (Prozentränge und IQ; Vertrauensintervall) durchgeführt. Die Ergebnisse werden im Anschluss an die Testdurchführung sowohl in Form einer Tabelle als auch in Form eines Profils im Ergebnisausdruck dargestellt. Eine Profilanalyse nach der Methode der psychometrischen Einzelfalldiagnostik liefert Hinweise auf diagnostisch abgesicherte individuelle Stärken und Schwächen der Testperson. Darüber hinaus bietet INSBAT-2 auch die Möglichkeit einer automatischen Übertragung der Testresultate in eine Befundvorlage. Informationen zum Testverlauf der einzelnen Subtests können den Testprotokollen entnommen werden.

## Zuverlässigkeit

Aufgrund der adaptiven Vorgabe der Subtests und der dafür notwendigen Geltung des 1PL Rasch Modells kann jede gewünschte Messgenauigkeit vorab als Abbruchkriterium eingestellt werden. Als Default-Einstellung wurde für die einzelnen Subtest eine SEM gewählt, der einer Reliabilität von  $r=0.70$  entspricht. Die zeitliche Stabilität der Testwerte bei einem Zeitintervall von rund 10 Monaten schwankt zwischen 0.70 und 0.90. Für die randomisierte Vorgabe konnte mittels einer Simulationsstudien gezeigt werden, dass im Mittel eine Reliabilität von 0.7 erzielt werden kann.

## Gültigkeit

Befunde zur Passung des 1PL Rasch Modells belegen die Annahme, dass mit den einzelnen Subtests jeweils eine latente Fähigkeitsdimension erfasst wird. Zudem zeigte sich auch, dass Personen unterschiedlicher soziodemografischer Gruppen (z. B.: Alter, Geschlecht, Bildung etc.) weder systematisch bevorzugt, noch benachteiligt werden. Die Ergebnisse in den einzelnen Subtests hängen somit ausschließlich von der zu messenden Fähigkeit der Testpersonen ab. Für die Praxis ergibt sich hieraus der Vorteil der Verrechnungsfairness und -sicherheit.

Die Aufgaben der einzelnen Subtests wurden mit Hilfe der Automatisierten Itemgenerierung auf Basis aktueller kognitions- und neurowissenschaftlicher Modelle konstruiert, aus denen die kognitiven Prozesse hervorgehen, die für die jeweils interessierende Fähigkeit typisch und relevant sind. In Studien mit dem Linear Logistischen Testmodell (LLTM) konnte gezeigt

werden, dass die Schwierigkeit der Aufgaben der Itempools zu den einzelnen Subtests gut durch Itemmerkmale vorhergesagt werden können, die diese kognitiven Prozesse ansprechen. Die Korrelation zwischen den anhand der Itemmerkmale prognostizierten Itemschwierigkeiten und den empirisch ermittelten Itemschwierigkeiten liegen zwischen  $R=0.86$  und  $R=0.96$ . Dies bedeutet, dass die theoretisch relevante Itemmerkmale zwischen 64 Prozent und 88 Prozent der Unterschiede in den Itemschwierigkeiten der 12 Subtests erklären. Da im 1PL Rasch Modell die Item- und Personenparameter auf einer gemeinsamen Skala gemessen werden, lässt sich aus diesem Befund ableiten, dass 64 Prozent bis 88 Prozent der Unterschiede in den Personenparametern der Testpersonen durch Unterschiede in den kognitiven Prozessen erklärbar sind, die für die jeweils gemessene Fähigkeit charakteristisch sind. Für die Mehrheit der Subtests der INSBAT-2 konnte dieser Befund auch in unabhängigen Studien mit anderen Itempools repliziert werden. Embretson (1983) spricht in diesem Zusammenhang auch von einem Nachweis der Konstruktrepräsentation. Für die Praxis ergibt sich hieraus der Vorteil der Interpretationseindeutigkeit. Diese Befunde können zudem auch als empirische Belege für die inhaltlich-logische Validität der Subtests angesehen werden. Befunde zur Faktorenstruktur der INSBAT-2 mit Hilfe von konfirmatorischen Faktorenanalysen sprechen für die Annahme, dass sich die Interkorrelationen zwischen den 12 Subtests durch die fünf ausgewählten Sekundärfaktoren erklären lassen. Die fünf Sekundärfaktoren sind zwar miteinander korreliert, jedoch auch klar voneinander separierbar. Dies spricht für die konvergente und diskriminante Validität der Subtests der INSBAT-2. Die vorliegenden Ergebnisse sprechen zudem auch für die Annahme, dass sich die Interkorrelationen der Subtests der INSBAT-2 am besten durch fünf Sekundärfaktoren und einem Generalfaktor der Intelligenz (G) beschreiben lassen. Dieses Modell weist nicht nur eine gute Passung auf die vorliegenden Daten auf, sondern beschreibt diese auch besser als eine Reihe alternativer Faktorenmodelle. Zusammenfassend betrachtet sprechen diese Befunde somit für die Konstruktvalidität der INSBAT-2. Die vorliegenden Befunde zur Konstruktvalidität werden zudem durch eine Reihe unabhängiger Studien abgerundet, in denen Subtests der INSBAT-2 zusammen mit Subtests aus anderen Intelligenztests vorgegeben wurden. Hinweise auf die Kriteriumsvalidität liegen aus dem Bereich der Flugpsychologie (Auswahl von Pilotenanwärtern) und der Bildungsberatung (Prognose des Ausbildungserfolgs an Fachhochschulen, Hochschulen und Pädagogischen Akademien) vor. Zudem erwies sich INSBAT-2 auch als valide für die Vorhersage beruflicher kreativer/innovativer Leistungen in unterschiedlichen Bereichen. Darüber hinaus erfüllt INSBAT-2 aufgrund der theoriegeleiteten Konstruktion und den vorliegenden empirischen Befunden zu ihren Gütekriterien zudem empirisch nachweisbar alle Voraussetzungen für die Generalisierung von Belegen zur prognostischen Kriteriumsvalidität aus aktuellen Metaanalysen im deutschsprachigen und internationalen Raum.

### Normen

Es stehen Normen einer nach Alter und Geschlecht repräsentativen Normstichprobe im Umfang von 714 Personen im Alter von 14 bis 70 Jahren zur Verfügung. Neben den Gesamtstichproben liegen auch Teilstichproben getrennt nach Geschlecht, Alter, und Bildung vor.

### Durchführungsdauer

Abhängig von der Auswahl der Subtests bzw. der vorab definierten Zielreliabilität.