

Glossar zu den Studienbriefen „Individualdiagnostik“, „Vergleichsarbeiten“ und „Bildungsmonitoring auf der Systemebene“

Alpha-Koeffizient

s. interne Konsistenz.

Standardsetzung

Hier: Inhaltliche Festlegung der in Bildungsstandards formulierten Kompetenzanforderungen

Abweichung

s. signifikante Abweichung.

Adjustierung der Testleistung

Berücksichtigung von Hintergrundinformationen der Schülerinnen und Schüler, die einen „fairen“ Vergleich von Leistungswerten ermöglichen sollen.

Akzeptanz

Akzeptanz in der Lehrerschaft gilt als wichtige Voraussetzung für den Erfolg von Vergleichsarbeiten.

Allgemeine mathematische Kompetenzen

Eine der drei Dimensionen der Bildungsstandards im Fach Mathematik. Sie bezieht sich auf die Art und Weise, wie Mathematik betrieben wird. Bei VERA 3 werden z.B. die folgenden sechs Kompetenzen unterschieden: „Technische Grundfertigkeiten“, „Problemlösen“, „Kommunizieren“, „Argumentieren“, „Modellieren“ und „Darstellen“

Anforderungsbereiche (in den Bildungsstandards Mathematik)

In den *Anforderungsbereichen* wird die kognitive Komplexität einer Aufgabe berücksichtigt. Differenziert werden drei nach zunehmender Komplexität geordnete Bereiche: „Reproduzieren“, „Zusammenhänge herstellen“ und „Verallgemeinern und Reflektieren“.

Aufgabe

Inhaltliche Einheit innerhalb eines Tests. Eine Aufgabe kann mehrere Items enthalten.

Aufgabenauswahl

s. Aufgabenkonstruktion/-auswahl bei Vergleichsarbeiten.

Aufgabenbeispiele

Hier: (Test-)Aufgaben, welche die in den Lehrplänen formulierten Anforderungen konkretisieren. Beispielaufgaben der Vergleichsarbeiten sind auf der Internetseite des IQB verfügbar (<http://www.iqb.hu-berlin.de/vera>).

Aufgabenerstellung

s. Aufgabenkonstruktion/-auswahl bei Vergleichsarbeiten.

Aufgabenformat true – false - not given

Mehrfach-Wahl-Aufgabe mit den Alternativen „wahr“ – „falsch“ – „keine Information im Text“.

Aufgabenkonstruktion/-auswahl bei Vergleichsarbeiten

Testaufgaben, die bei Vergleichsarbeiten eingesetzt werden sollen, durchlaufen mehrere Phasen der Qualitätsprüfung. Darunter fallen die inhaltliche (fachliche) Prüfung der Validität, die Überprüfung der Objektivität mit Hilfe von Zweitkodierungen, die Überprüfung der Reliabilität und ggf. die Überprüfung der Modellpassung in Bezug auf ein Testmodell der probabailistischen Testtheorie, beispielsweise dem Rasch-Modell.

Auswertungsobjektivität

Auswertungsobjektivität liegt vor, wenn die Punkteverteilung für die Testantworten nicht vom Testleiter abhängt. Bei der Auswertung eines Schulleistungstest sollten verschiedene Lehrkräfte daher zum selben Ergebnis kommen. Die Auswertungsobjektivität eines Tests ist in der Regel gegeben, wenn die richtige bzw. falsche Beantwortung eines Items eindeutig definiert ist.

Begleitmaterial

Zur didaktischen Unterstützung von Vergleichsarbeiten werden unter anderem Anregungen zur Nutzung der Testergebnisse und –materialien, aber auch die Ergebnisse der jeweiligen Schule, in speziell hierfür durch die Länder eingerichteten und durch ein Passwort geschützten Internetportalen bereitgestellt.

Beispielaufgaben

s. Aufgabenbeispiele.

Belastungsgruppe

s. Standortgruppe.

Benchmarks

Benchmarks beschreiben im Kontext des Bildungsmonitorings Leistungsergebnisse sehr erfolgreicher Bildungssysteme im In- und Ausland. Mit diesen Leistungsergebnissen kann das ‚eigene‘ Ergebnis verglichen werden.

Bezugsgruppen

In großangelegten Studien ist es nicht möglich, eine Grundgesamtheit von Schülerinnen und Schülern zu testen. Um Vergleiche ziehen zu können, muss deutlich gemacht werden, wie sich eine Stichprobe zusammensetzt. Bezugsgruppen im Bildungsmonitoring sind nicht mit Bezugsnormen (siehe unten) zu verwechseln.

Bezugsnormen

Bezugsnormen sind Vergleichswerte, welche die individuelle Merkmalsausprägung in Relation zu a) entwicklungsbedingten Veränderungen individueller Merkmalsausprägung im Laufe der Zeit, b) der Merkmalsausprägung einer entsprechenden Referenzgruppe oder c) einem vorgegebene Kriterium setzen.

Big fish little Pond

„Big fish little Pond“ (Fischteicheffekt): In einer Klasse leistungsschwächerer Mitschüler (little pond) entwickelt eine SchülerIn eine stärkere Lernmotivation als in einer Klasse leistungsstärkerer Mitschüler (big pond), da ihre/seine Leistungen stärker herausragen und sie/er diese herausragende Position zu halten gewillt ist. Unter bestimmten Bedingungen (z.B. Lerngruppe mit herausragendem Renomee) kann dieser Effekt aufgehoben oder sogar ins Gegenteil („Reflected-Glory-Effect“) verkehrt werden.

Bildungsadministration

Alle Beteiligten der Verwaltung, die mit dem Bereich Bildung beschäftigt sind, z.B. Schulämter.

Bildungsberichterstattung

Kontinuierliche Berichterstattung über den Bildungsbereich.

Bildungsmonitoring

Die systematische und auf Dauer angelegte Beschaffung und Aufbereitung von Informationen über ein Bildungssystem und dessen Umfeld. Bildungsmonitoring dient als Grundlage für die Bildungsplanung und bildungspolitische Entscheidungen.

Bildungsstandards

s. nationale Bildungsstandards.

BLK

Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung.

BLK-Projekt „SINUS-Transfer“

s. Projekt „SINUS-Transfer“.

Boxplot

Boxplots werden in der Statistik verwendet, um die Verteilung von Daten grafisch darzustellen. Einem Boxplot lassen sich verschiedene statistische Informationen entnehmen, z. B. Minimum/Maximum der Datenreihe, Median und Spannweite.

can do statements

Aussagen, die beschreiben, was ein Schüler kann. Dabei ist besonders wichtig, dass sie immer positiv formuliert sind.

Cronbachs Alpha

Cronbachs Alpha dient als ein Maß der internen Konsistenz einer aus mehreren Items zusammengesetzten Skala. Eine interne Konsistenz kann als Maß für die durchschnittliche Korrelation aller Items einer Skala verstanden werden und stellt damit ein Maß der Reliabilität einer Skala dar.

Curriculare Gültigkeit

s. curriculare Validität.

Curriculare Orientierung

Eine curriculare Orientierung in einer Untersuchung besteht, wenn das Curriculum als Grundlage für die Testerstellung verwendet wird und überprüft werden soll, ob die Inhalte des Curriculums erreicht werden. Gegenteil: Literacy-Orientierung.

Curriculare Validität

Übereinstimmung von Testaufgaben mit Lehrplänen an getesteten Schulen.

Daten

Hier: Material, das (a) während der Evaluation gesammelt wird, um zuverlässige Informationen über den untersuchten Arbeitsbereich oder Gegenstand zu erhalten oder (b) der Schulverwaltung bereits zu statistischen Zwecken vorliegt.

Datenschutz

Hier: Schutz schul- und personenbezogener Daten der Schülerinnen und Schüler vor dem Missbrauch durch Weitergabe an Dritte.

Delfin 4

„Diagnostik, Elternarbeit, Förderung der Sprachkompetenz in Nordrhein-Westfalen bei 4-Jährigen“ (Delfin 4) ist ein Projekt der Sprachförderung in Nordrhein-Westfalen. Zwei Jahre vor Einschulung werden alle Kinder getestet, anschließend wird eine Sprachförderung durchgeführt.

DESI (Deutsch Englisch Schülerleistungen International)

In Deutschland, Österreich und der Schweiz durchgeführte Schulleistungsstudie. Zu zwei verschiedenen Zeitpunkten (September/Oktober 2003 und Mai/Juni 2004) wurden die Kompetenzen von Schülerinnen und Schülern neunter Klassen in unterschiedlichen Schulformen in den Bereichen Deutsch und Englisch erfasst. Die repräsentative Studie testete dabei rund 11.000 Schülerinnen und Schülern.

Deskriptive Statistik

In der deskriptiven Statistik (auch beschreibende Statistik oder empirische Statistik genannt) werden vorliegende Daten in einer geeigneten Weise beschrieben, aufbereitet und zusammengefasst.

Determinanten von Schulleistungen

Als individuelle Determinanten (bzw. Korrelate) von Schulleistungen gelten kognitive Fähigkeiten (z.B. Intelligenz), motivationale Bedingungen (z.B. das Fähigkeitsselbstkonzept) und sozioökonomische Variablen. Darüber hinaus spielt die Qualität des Unterrichts eine zentrale Rolle für die Erklärung von Schulleistungen.

Deutsche Kultusministerkonferenz (KMK)

Die Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland ist ein freiwilliger Zusammenschluss der für Bildung, Erziehung und Forschung sowie kulturelle Angelegenheiten zuständigen Minister bzw. Senatoren der Länder. Die Kultusministerkonferenz behandelt nach ihrer Geschäftsordnung „Angelegenheiten der Kulturpolitik von überregionaler Bedeutung mit dem Ziel einer gemeinsamen Meinungs- und Willensbildung und der Vertretung gemeinsamer Anliegen“.

Diagnosearbeit

s. flächendeckende Vergleichsarbeit.

Diagnostik

Bei pädagogisch-psychologischen Diagnostik steht die Fundierung diagnostischer Entscheidungen in den verschiedensten Bereichen pädagogischer und psychologischer Tätigkeiten (z.B. Selektions- und Platzierungsentscheidungen) im Vordergrund (vgl. Leutner, 2006).

Diagnostische Instrumente

Unter diagnostischen (Test)-Instrumenten sind im allgemeinen Sinne Werkzeuge zu verstehen, mit denen beobachtbare Sachverhalte bestimmten Kategorien genau zugeordnet und interpretiert werden können.

Diagnostische Kompetenz von Lehrkräften

Diagnostische Kompetenzen von Lehrerinnen und Lehrern werden als Fähigkeit dieser verstanden, Schülerinnen und Schüler hinsichtlich der Ausprägung bestimmter Merkmale treffend zu beurteilen. Im weiteren Sinne wird unter diagnostischer Kompetenz auch die Fähigkeit von Lehrpersonen verstanden, die Schwierigkeit von Aufgaben für eine bestimmte Lerngruppe korrekt einzuschätzen (vgl. Schrader, 2001; Spinath, 2005).

Dichotomes Antwortformat

Auswertungsformat, bei dem Items entweder als richtig oder als falsch bewertet werden müssen.

Diskriminante Validität

Die diskriminante Validität gibt Auskunft darüber, wie sicher konstruktferne Kriterien erfasst werden.

Durchführungsobjektivität

Durchführungsobjektivität liegt vor, wenn der Testverlauf vom Testleiter unabhängig ist.

Ergänzungsaufgaben

s. halboffene Formate.

Ergänzungsstudien

Um intranationale Vergleiche durchführen zu können werden innerhalb der Studien des Bildungsmonitorings Ergänzungsstudien durchgeführt. Hier können zusätzliche Themenbereiche oder beispielsweise sozioökonomische Hintergründe der Schülerinnen und Schüler erfragt werden. Die Studien werden nach der internationalen Durchführung durchgeführt.

Ergebnisrückmeldungen

Ergebnisrückmeldungen aus Vergleichsarbeiten sollen Lehrpersonen Orientierung geben, inwieweit die erwarteten Standards innerhalb einer Klasse und innerhalb einer Jahrgangsstufe erreicht werden konnten.

Erste Fremdsprache

Die erste Fremdsprache (in der Regel Englisch, seltener Französisch) ergänzt die Fächer Deutsch und Mathematik bei den Vergleichsarbeiten in der 8. Jahrgangsstufe (VERA 8).

Erstsprache

Erstsprache beschreibt die Sprache, die ein Kind als erstes lernt bzw. am besten beherrscht.

Erwartungswert

Der Erwartungswert ist ein Begriff der Wahrscheinlichkeitsrechnung und beschreibt jenen Wert, der sich (in der Regel) bei oftmaligem Wiederholen des zugrunde liegenden Experiments als Mittelwert der Ergebnisse ergibt.

Evaluation

Sammlung, Verarbeitung und Interpretation von Informationen über schulische Arbeit. Sie hat das Ziel, zu gesicherten Beschreibungen zu kommen, Bewertungen nach klaren Kriterien durchzuführen und Entscheidungen über die Weiterentwicklung dieser Arbeit zu treffen.

Externe Evaluation

Externe Evaluation bedingt, dass die Verantwortung für die Gestaltung und Durchführung einer Evaluation außerhalb der Schule liegt und von Personen durchgeführt wird, die nicht unmittelbar für die Arbeit verantwortlich sind.

Fachdidaktik

In Abgrenzung zur (allgemeinen) Lehr-Lernforschung stehen in den Fachdidaktiken fach- bzw. domänenspezifische Lernprozesse in Forschung, Lehre und Entwicklung im Vordergrund.

Fähigkeitsniveau

Im Gegensatz zur Bezeichnung „Kompetenzniveaus“ wurde bei den Vergleichsarbeiten in der dritten Jahrgangsstufe (VERA 3) bis 2009 der Begriff „Fähigkeitsniveaus“ verwendet. Inzwischen werden bei VERA 3 ebenfalls fünf „Kompetenzniveaus“ unterschieden.

Fairer Vergleich

Vergleich der Ergebnisse von Schülerinnen und Schüler unter Berücksichtigung schulleistungsrelevanter Hintergrundmerkmale (siehe auch Determinanten von Schulleistungen).

Fairness

Fairness zählt zu den Nebengütekriterien und liegt dann vor, wenn Testpersonen nicht aufgrund ihrer ethischen, soziokulturellen oder geschlechtsspezifischen Gruppenzugehörigkeit diskriminiert werden.

Flächendeckende Vergleichsarbeit

Landesweit durchgeführte Vergleichsarbeit (s. Vergleichsarbeit).

Fluide Intelligenz

Fluide Intelligenz bezeichnet bildungsunabhängige, intellektuelle Fähigkeiten (bspw. Fähigkeit komplexe Beziehungen in neuartigen Situationen wahrzunehmen und zu erfassen).

Formate

s. offene, halboffene oder geschlossene (Aufgaben-) Formate.

FörMig

Das Projekt „Förderung von Kindern und Jugendlichen mit Migrationshintergrund“ (FörMig) wurde von der Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung (BLK) als große Modellinitiative als Reaktion auf das schlechte Abschneiden aller Schülerinnen und Schüler im Bereich der Lesekompetenz in PISA 2000 gestartet. Innerhalb des Projektes sollen innovative Projekte der Sprachförderung entwickelt und an andere weitergegeben werden.

Fremdsprache:

s. erste Fremdsprache.

Ganztagsschule, offen

Die Teilnahme am ganztägigen Angebot ist freiwillig, die Eltern der Schülerinnen und Schüler müssen die Teilnahme allerdings für mindestens ein Jahr als verbindlich erklären, die Schülerinnen und Schüler können sich an mindestens drei Tagen der Woche täglich mindestens sieben Zeitstunden in der Schule aufhalten und an einem Bildungs- und Betreuungsangebot teilnehmen.

Ganztagsschule, teilweise gebunden

Ein Teil der Schülerinnen und Schüler ist dazu verpflichtet, an mindestens drei Tagen der Woche für jeweils mindestens sieben Zeitstunden an den ganztägigen Angeboten der Schule teilzunehmen.

Ganztagsschule, voll gebunden

Alle Schülerinnen und Schüler sind verpflichtet, an mindestens drei Tagen der Woche für jeweils mindestens sieben Zeitstunden an den ganztägigen Angeboten der Schule teilzunehmen.

Gemeinsamer europäischer Referenzrahmen

Der GeR stellt eine Empfehlung für Fremdsprachenlernende und –lehrende dar. Er trägt so dazu bei, Spracherwerb und –kompetenz transparent und vergleichbar zu gestalten (s. auch GER-Niveau).

Genauigkeit

s. Reliabilität.

GeR

s. gemeinsamer europäischer Referenzrahmen.

GER-Niveau

Sechs verschiedene Stufen (A1, A2, B1, B2, C1, C2) des Gemeinsamen Europäischen Referenzrahmens für Sprachen lassen eine gemeinsame Niveauspezifikation zu.

Gesamtstrategie zum Bildungsmonitoring

Beschluss der Kultusministerkonferenz im Juni 2006 zur systematischen und wissenschaftlich abgesicherten Feststellung von Ergebnissen des Bildungssystems. Auf dieser Grundlage sollen mögliche Gründe für eventuell unbefriedigende Ergebnisse analysiert und daraus geeignete Reformmaßnahmen durch die Bildungspolitik und -verwaltung abgeleitet werden. Die Gesamtstrategie der Kultusministerkonferenz umfasst drei Säulen: Internationale Schulleistungsuntersuchungen, zentrale Überprüfung des Erreichens der Bildungsstandards im Ländervergleich (nationaler Vergleich), Vergleichsarbeiten zur landesweiten Überprüfung der Leistungsfähigkeit einzelner Schulen (regionaler Vergleich). Hinzu kommt die gemeinsame Bildungsberichterstattung von Bund und Ländern.

Geschlossene (Aufgaben-) Formate

Antwortformate, bei denen bereits Antwortalternativen vorgegeben sind. Als Beispiele seien Mehrfachwahlaufgaben oder Zuordnungsaufgaben genannt.

Grundbildungsaspekte

Bestandteile dieser sind: Lesekompetenz, mathematische Grundbildung und naturwissenschaftliche Grundbildung. Dabei wird nicht reines Faktenwissen, sondern die Anwendung der Fakten beachtet. Beispielsweise PISA untersucht Grundbildungsaspekte.

Grundintelligenz oder G-Faktor

Die Grundintelligenz (G-Faktor) repräsentiert Fähigkeiten, welche von verschiedenen Tests gleichermaßen gemessen werden. Der Begriff beinhaltet, dass in verschiedenen Intelligenzleistungen ein allgemeiner Intelligenzfaktor (G-Faktor) mitwirkt. Er ist Gegenstand von Spearman's Zwei-Faktoretheorie.

Halboffene (Aufgaben-) Formate

Halboffene (Aufgaben-) Formate sind beispielsweise durch Schlüsselworte, Symbole, ggf. auch zeichnerische Darstellung vollendete Aufgaben. Hierzu gehören unter anderem Ergänzungsaufgaben und Lückentexte.

Hamburger Schulleistungstest (HST)

Der Hamburger Schulleistungstest (HST) soll die wesentlichen Aspekte schulischen Lernens zuverlässig bilanzieren. Um die entsprechende Validität zu gewährleisten, wurde der HST laut Testbeschreibung unter Berücksichtigung der curricularen Anforderungen der Klassen 4 und 5 konstruiert. Er untersucht fünf Inhaltsbereiche (Sprachverständnis, Leseverständnis, Rechtschreibung, Informationsentnahme aus Karten, Tabellen und Diagrammen und Mathematik).

Hintergrundmerkmale

s. Kontextmerkmale.

Hochbegabung

Hochbegabung wird als individuelles Fähigkeitspotential für außergewöhnliche Leistungen in einem bestimmten Bereich (z.B. Mathematik, Fremdsprachen, Musik, Kunst, Sport) oder in mehreren Bereichen (sog. Multitalente oder Universalbegabungen) definiert. Für die Hochbegabungsentwicklung bzw. die Umsetzung der Fähigkeitspotentiale in herausragende Verhaltensleistungen sind neben kognitiven und motivationalen Lernvoraussetzungen auch soziale Lernumweltbedingungen im Sinne kreativer Lernumwelten bedeutsam.

Hörverstehen

Einer der bisher operationalisierten Kompetenzbereiche im Fach Englisch bei den Vergleichsarbeiten in der achten Jahrgangsstufe (VERA 8).

ICC

s. Itemcharakteristische Kurve.

IGLU-Erhebung (Internationale Grundschul-Lese-Untersuchung)

Erstmals 2001 von 35 Staaten mit insgesamt rund 147.000 Schülern durchgeführte internationale Grundschul-Lese-Untersuchung. In Deutschland nahmen alle 16 Bundesländer mit mehr als 10.000 Schüler und Schülerinnen teil.

Indikatoren

Als ein Indikator wird ein (messbarer) Sachverhalt gesehen, der auf ein ausgewähltes Phänomen bezogen Aussagekraft besitzt.

Individualdiagnostik

Individualdiagnostik dient der Vorbereitung von Entscheidungen im Einzelfall, z.B. dem Aufstellen von individuellen Förderplänen.

Individuelle Bezugsnorm

Individuelle Bezugsnorm ist der Bezug der aktuellen individuellen Merkmalsausprägungen auf vorherigen Ausprägungen.

Inhaltliche Beschreibung von Kompetenzniveaus

Um Kompetenzniveaus inhaltlich nutzbar zu machen, untersuchen Expertinnen und Experten Aufgabencluster in Bezug auf gemeinsame Anforderungsmerkmale. Sie schließen dabei von gemeinsamen Anforderungsmerkmalen auf lösungsrelevante Kompetenzen. Lehrerinnen und Lehrer erhalten dadurch einen anschaulichen Eindruck davon, welchen Anforderungen ihre Schülerinnen und Schüler im relevanten Kompetenzbereich gerecht werden bzw. gerecht werden sollten.

Inhaltsbezogene mathematische Kompetenzen

s. Mathematische Leitideen

Inhaltsvalidität

Inhaltsvalidität gibt Auskunft inwieweit die Items das zu erfassende Zielkonstrukt repräsentieren. Ziel ist es bei der Konstruktion von Test-Items eine repräsentative Auswahl aus der Grundgesamtheit der Items darzustellen, welche für die Erfassung des Zielkonstrukts in Frage kommen.

Input-Orientierung

Nach der Input-Orientierung werden Input-Variablen wie beispielsweise Qualität der Lehrerbildung, richtige Rahmenbedingungen für die Autonomie der einzelnen Schule, Organisation der Schulaufsicht, zweckmäßiger Einsatz der finanziellen Mittel und Qualität der Lehrpläne untersucht und die relevanten Mittel in ausreichendem Umfang den Schulen zur Verfügung gestellt.

Input-Steuerung

Art der Steuerung des Bildungswesens. Vorgegeben wird, beispielsweise durch Lehrpläne, welchen Input die Schülerinnen und Schüler erhalten sollen.

Institut zur Qualitätsentwicklung im Bildungswesen

Das Institut zur Qualitätsentwicklung im Bildungswesen (IQB) ist eine wissenschaftliche Einrichtung der deutschen Bundesländer mit Sitz in Berlin. Im Zentrum der Arbeit des IQB steht die Normierung und Überprüfung von Bildungsstandards. Seit 2009 ist das IQB bundesweit für die Erstellung von Aufgaben für Vergleichsarbeiten verantwortlich.

Intelligenz

In der Psychologie ist Intelligenz ein Sammelbegriff für die kognitive Leistungsfähigkeit des Menschen. Beispiele sind die Fähigkeiten, zu verstehen, zu abstrahieren, Probleme zu lösen, Wissen anzuwenden und Sprache zu verwenden. Exemplarisch seien die Generalfaktorthorie von Spearman (1904) und das Modell der fluiden und der kristallinen Intelligenz von Catell (1957) angeführt.

Inter-Item-Korrelation

s. interne Konsistenz.

Internationaler Vergleich

Hierzu gehören internationale Vergleichsstudien wie IGLU/PIRLS, TIMSS, PISA, DESI, die im mehrjährigen Rhythmus erhoben werden.

Interne Evaluation

Interne Evaluation bedingt, dass die Verantwortung für die Gestaltung und Durchführung einer Evaluation in der Schule liegt und von den unmittelbar Verantwortlichen selbst durchgeführt wird.

Interne Konsistenz

Interne Konsistenz geht davon aus, dass jedes Item im Wesentlichen als Paralleltest behandelt werden kann. Ein Zusammenhang (Korrelation) zwischen den Items spricht für die Homogenität einer Skala. Die gebräuchlichste Methode zur Berechnung der internen Konsistenz ist der von Cronbach (1951) entwickelte „Alpha –Koeffizient“. Im Allgemeinen wird ein Alpha-Koeffizient von 0.66 oder besser als akzeptabel bewertet.

Internetportal

s. Passwort-geschütztes Internetportal.

Interpretationsobjektivität

Interpretationsobjektivität liegt vor, wenn sich die Interpretation eines Testresultats an Vergleichswerten (Normen) orientiert, welche als Maßstab dienen. Als Vergleichswerte dienen die Testergebnisse repräsentativer Stichproben. Standardisierte Testverfahren sollten daher entsprechende Vergleichsnormen (Schulformbezogene Normen, Altersnormen, Bildungsnormen) in tabellarischer Form aufweisen.

Intervallskala

Skala, bei der Differenzen zwischen den Punktwerten identische Differenzen zwischen der Ausprägung des betrachteten Merkmals bedeuteten.

Intranational

Innerhalb eines Landes, beispielsweise: Vergleich aller deutschen Bundesländer.

IQB

s. Institut zur Qualitätsentwicklung im Bildungswesen.

IRT-Modell

s. probabilistische Testtheorie.

Item

Kleinste, bewertbare Einheit eines Tests. Ein oder mehrere Items bilden eine Aufgabe.

Item Response Theory (IRT)

s. probabilistische Testtheorie.

Itemcharakteristische Kurve (ICC)

Die ICC gibt für Test-Items die Wahrscheinlichkeit einer richtigen Lösung in Abhängigkeit der Schülerkompetenz) an.

Itemstichprobe

Die komplette Anzahl aller möglichen Items eines Inhaltsbereiches kann Schülerinnen und Schülern aus ökonomischen Gründen in aller Regel nicht vorgelegt werden, sodass man bei der Testentwicklung bestrebt ist, für einen bestimmten Inhaltsbereich eine möglichst repräsentative Stichprobe von Items auszuwählen.

Jahrgangsstufentest

s. flächendeckende Vergleichsarbeit.

Kernlehrpläne

Kernlehrpläne orientieren sich an den Bildungsstandards der Kultusministerkonferenz (KMK) und konkretisieren diese. Sie beschreiben die Anforderungen, die am Ende einer Klassenstufe erreicht werden sollen in Form von Kompetenzen. Die Kernlehrpläne können somit als Umsetzung der Bildungsstandards der KMK auf der curricularen Ebene verstanden werden.

Klasseninterne Vergleichsmaßstäbe

Ein Vergleichsmaßstab beruhend auf einer sozialen Bezugsnorm, der den Förderbedarf einer entsprechend leistungsschwachen Referenzgruppe (Klasse oder Kurs) verbergen kann.

Klassische Testtheorie

Die Klassische Testtheorie besagt in ihrer Grundannahme, dass der beobachtete Testwert aus dem wahren Wert und einem unsystematischen Messfehler besteht.

KMK

s. Deutsche Kultusministerkonferenz.

Kompetenz

Unter Kompetenzen versteht man die bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen. Im Unterschied zur Intelligenz sind Kompetenzen in deutlich stärkerem Maße lern- und trainierbar (nach Weinert, 2001; vgl. auch Klieme & Leutner, 2006; Köller, 2009).

Kompetenzbereich

Bei Vergleichsarbeiten werden jeweils bestimmte Teilleistungsbereiche eines Unterrichtsfachs mit dem Ziel operationalisiert, die Kompetenzen der Schülerinnen und Schüler möglichst reliabel und valide zu erfassen.

Kompetenzerwartung / Kompetenzanforderung

In den Lehrplänen festgelegter, konkreter Anforderungskatalog an zu erreichenden Kompetenzen von Schülerinnen und Schüler.

Kompetenzmodell

Modell, welches die relevanten Leistungsanforderungen in Bezug auf einen bestimmten Inhaltsbereich definiert. Unterschieden wird ggf. zwischen einem Kompetenzstrukturmodell, welches die Anforderungen inhaltlich gliedert und einem Kompetenzentwicklungsmodell, welches Stufen des Erwerbs entsprechender Anforderungen beschreibt.

Kompetenzniveau

Konsequente Bereiche der Kompetenzskala werden auf Basis von inhaltlichen Überlegungen und unter Berücksichtigung der Itemschwierigkeiten als Kompetenzniveaus ausgewiesen. Jedes Niveau wird durch bestimmte kognitive Prozesse und Handlungen von bestimmter Qualität charakterisiert. Kompetenzniveaus ermöglichen eine kriteriumsorientierte Interpretation der Ergebnisse.

Kompetenzskala

s. Kompetenzmodell.

Kompetenzstufe

s. Kompetenzniveau.

Kompetenztests

Seit dem Schuljahr 2002/2003 in Thüringen unter dem Namen Thüringer Kompetenztests landesweit durchgeführte Leistungstests in den Fächern Deutsch und Mathematik in der Klassenstufe 3 sowie Mathematik, Deutsch und Englisch oder Französisch in der Klassenstufe 8 (ursprünglich Klassenstufe 6).

Konstanzer Beschlüsse

Beschlüsse der KMK, die die regelmäßige Teilnahme an weiteren internationalen Studien beinhalten.

Konstruktvalidität

Konstruktvalidität wird überprüft, indem Hypothesen über das zu erfassende Konstrukt und die Zusammenhänge zu verwandten Konstrukten formuliert werden. Es kann zwischen konvergenter und diskriminanter Validität unterschieden werden.

Kontextgruppe

s. Standortgruppe.

Kontextmerkmale

Merkmale der Schülerinnen und Schüler, die sich dem Einfluss von Schule und Unterricht entziehen. Hierunter fallen individuelle kognitive Eingangsvoraussetzungen, aber auch sozioökonomische Merkmale der Schülerschaft.

Kontrollgruppe

Gruppe von Schülerinnen und Schüler mit ähnlichen Voraussetzungen wie die untersuchte Gruppe. Dabei erhalten beide Gruppen beispielsweise einen anderen Unterricht, der Unterschied der Kontrollgruppe zur eigentlich untersuchten Gruppe wird analysiert.

Konvergente Validität

Konvergente Validität gibt Auskunft darüber, wie stark der Zusammenhang des untersuchten Testinstrumentes mit konstruktnahen Kriterien ausfällt.

Konzeptwissen

Wissen über die Wechselbeziehungen zwischen den einzelnen Elementen des Basiswissens innerhalb eines größeren Zusammenhangs, das ein gemeinsames Funktionieren sichert.

Korrelation

Maß der Stärke eines Zusammenhangs zwischen zwei (metrischen) Variablen. Je höher der die gemeinsame Streuung beider Variablen, desto höher die Korrelation. Korrelationen können zwischen $r=-1$ und $r=1$ liegen; ein Betrag von $r=0$ bedeutet keinen Zusammenhang, hingegen ein Betrag von $r=|1|$ einen perfekten Zusammenhang.

Kristalline Intelligenz

Kristalline Intelligenz bezieht sich auf die umweltbedingte, d.h. im Wesentlichen auf Lernerfahrungen (z.B. aus der Schule) beruhende Komponente von Intelligenz.

Kriteriale Bezugsnormen

Bezugsnorm, bei welcher der Vergleich der individuellen Merkmalsausprägung mit einem vorgegebenen Kriterium Berücksichtigung findet.

Kriteriale Perspektive (einer Lehrperson)

Die möglichst treffende Verortung des Leistungsstandes der Lerngruppe durch eine Lehrperson in Bezug auf definierte Standards der Lehrpläne. Diese kann z.B. bei Vergleichsarbeiten durch Überprüfung des tatsächlichen Lernstandes kritisch reflektiert werden.

Kriteriumsvalidität

Kriteriumsvalidität liegt vor, wenn vom Testergebnis erfolgreich auf ein Kriterium, also auf ein Verhalten außerhalb der Testsituation geschlossen werden kann.

Kulturelles Kapital

Mit dem Begriff des kulturellen Kapitals beschreibt Bourdieu (1983) körpergebundene und in der Familie weitergegebene Aspekte der Bildung, die sich in sozialen Beziehungen als nutzbringend erweisen. Hierunter fallen auch kultureller Güter.

Kumulativer Kompetenzerwerb

Systematisches und vernetztes Lernen.

Ländervergleichsgruppe

Vorgehen, das bei IGLU verwendet wurde. Länder mit ähnlichen Voraussetzungen werden gruppiert, so kann sich jedes Land mit Ländern vergleichen, die einen ähnlichen sozialen und wirtschaftlichen Hintergrund haben.

Längsschnittstudie

Längsschnittstudien beschreiben Studien, die eine Situation oder Kompetenz nicht nur zu einem Zeitpunkt untersuchen. In Längsschnittstudien werden die Testpersonen zu mindestens zwei Zeitpunkten getestet, sodass sich eine zeitliche Entwicklung darstellen lässt.

Large-Scale Assessments

Leistungsmessungen mit einer hohen Anzahl an Testpersonen.

Latente Merkmale

Als latente Merkmale werden nicht direkt erfassbare Merkmale verstanden; s. Merkmal.

LAU (Aspekte der Lernausgangslage und Lernentwicklung)

1995 von der Hamburger Schulbehörde initiierte Studie des Bildungsmonitorings, bei der die Lernausgangslage und Lernentwicklung aller Schüler untersucht wurde, die sich im Jahr 1996 in Hamburg in der fünften Klasse (LAU 5), 1998 in der siebten (LAU 7), 2000 in der neunten (LAU 9), 2002 in der elften (LAU 11) und 2005 in der 13. Klasse (LAU 13) befanden.

Lehrplanvalidität

s. curriculare Validität

Lehrzielorientierter Test

Schulleistungstest, welcher prüft, ob oder in welchem Ausmaß Schülerinnen und Schüler ein Lehrziel erreicht haben (Klauer, 1987).

Leistungsprüfsystem (LPS)

Leistungsprüfsystem (LPS) ist ein Intelligenztest, welcher auf Grundlage des Intelligenzmodells der sieben Primärfaktoren von Thurstone (1938) basiert und das das Begabungsprofil von Probanden zwischen 9 und 50 Jahren in den Bereichen Sprachverständnis, Wortflüssigkeit, numerische Fähigkeit, räumliche Vorstellung, Wahrnehmungsgeschwindigkeit, Gedächtnisleistung und die Fähigkeit des schlussfolgernden Denkens erfasst.

Leistungsstreuung

Die PISA-Erhebungen haben auf relativ deutliche Unterschiede zwischen den Schulformen in Deutschland hingewiesen. Gleichzeitig weist das deutsche Schulsystem eine breite Streuung erreichter Kompetenzanforderungen zwischen Schulen derselben Schulform aus. Um dieser Leistungsstreuung gerecht zu werden, können zum Beispiel unterschiedliche Testhefte verschiedener Schwierigkeitsspektren verwendet werden.

Leistungstests

Leistungstests dienen der Erfassung von Leistungsmerkmalen. Im Zentrum steht in der Regel die Lösung vorgegebener Problemstellungen. Ein typischer Anwendungsbereich ist die Schule, wo fachspezifische oder fächerübergreifende Leistungstests zum Einsatz kommen können.

lernstand 8

In NRW tragen Vergleichsarbeiten (Lernstandserhebung) in der achten Jahrgangsstufe den Namen *lernstand 8*.

Lernstandsdiagnose

Summe diagnostischer Maßnahmen zur Erfassung des Lernstands.

Lernstandserhebung

s. flächendeckende Vergleichsarbeit.

Lernziele

Lernziele sind auf die Steigerung der eigenen Kompetenzen ausgerichtet. Der Vergleich der eigenen Leistung über einen bestimmten Zeitraum hinweg dient dem Lerner als Maßstab zur Bewertung der Kompetenzsteigerung (individuelle Bezugsnorm).

Lesen

s. Leseverstehen.

Leseverstehen

In den sprachlichen Fächern Deutsch, Englisch und Französisch lassen sich vor allen kommunikative Kompetenzen sinnvoll erfassen. Einer der daraus resultierenden Kompetenzbereiche ist das Leseverstehen.

Literacy

s. auch Grundbildungsaspekte, Basiskompetenzen, die für das alltägliche Leben von wichtiger Bedeutung sind.

Literacy-Orientierung

Studien, die nicht Inhalte eines Curriculums abtesten, folgen einer literacy-Orientierung, sie testen Grundbildungskompetenzen der Schülerinnen und Schüler. Gegenteil: curriculare Orientierung.

Lösungswahrscheinlichkeit

In der klassischen Testtheorie die Anzahl korrekter Lösungen geteilt durch die Anzahl der dieses Item bearbeitenden Schülerinnen und Schüler. Im Rasch-Modell ergibt sich die

Lösungswahrscheinlichkeit aus dem Verhältnis von Item- Schwierigkeit und SchülerKompetenz.

Matching-Item

s. Zuordnungsaufgaben.

Mathematische Leitideen

Eine der Dimensionen der Bildungsstandards im Fach Mathematik. Beschrieben werden: Phänomene einer mathematischen Betrachtungsweise, die den Stoffgebieten der Schulmathematik (Algebra, Arithmetik, Geometrie und Stochastik) vorausgehen und auch konzeptionell von diesen abzugrenzen sind. Das Konzept der Leitideen greift Überlegungen der Freudenthal-Didaktik (Freudenthal, 1977/1983) auf.

Maximalstandards

Exzellente und nur unter sehr günstigen Umständen zu erbringende Leistungen, die die Regelerwartungen zu einem bestimmten Zeitpunkt der Bildungsbiographie weit übertreffen.

Mehrfächertests

Mehrfächertests sind Schulleistungstests, welche die Schulleistung in relevanten Fächern überprüfen.

Mindestbeteiligungsquote

Wird vor den Systemmonitoring-Studien festgelegt, beschreibt prozentual den Anteil der Schülerinnen und Schüler der Stichprobe, die sich tatsächlich an der Untersuchung beteiligen müssen.

Mindeststandards

Umfasst die Definition von Anforderungen, die alle Schülerinnen und Schüler bis zu einem bestimmten Abschnitt ihrer Bildungsbiographie erfüllen sollen.

Mittelwert

Kurzbezeichnung für den arithmetischen Mittelwert. Der Mittelwert setzt sich zusammen aus der Summe der Einzelwerte aller Fälle dividiert durch die Fallzahl. Umgangssprachlich wird für Mittelwert oft der Begriff „Durchschnittswert“ verwendet.

Mittelwertvergleich

Statistisch ermittelte Unterschiede können zufällig entstehen. Deshalb müssen beim Vergleich von Mittelwerten die Konfidenzintervalle berücksichtigt werden. Nur wenn sich die beiden Konfidenzintervalle zweier Mittelwerte nicht überschneiden kann man von einem signifikanten (bedeutsamen) Unterschied zwischen den beiden Mittelwerten sprechen. Daraus folgt, dass sich Schülerleistungen zwischen Staaten und Bundesländern relativ gut miteinander vergleichen lassen, da hier die Stichproben groß und die Konfidenzintervalle klein sind. Je niedriger das Aggregationsniveau der Schülerleistungen wird (z.B. Leistungen für einzelne Schulen oder einzelne Schulklassen), desto kleiner werden die Stichproben und desto größer werden die Konfidenzintervalle.

Motivation

Motivation entsteht, wenn Motive durch eine Situation angeregt und somit Verhaltenswirksam werden.

Multi-Matrix-Design

Design für die Zuordnung von Test-Items zu Personen, bei dem jede Person nur eine Auswahl der gesamten Menge aller verwendeten Test-Items vorgelegt bekommt. Durch Rotation der Items (Items werden mit jeweils anderen Items auf verschiedene Testhefte verteilt) können die erzielten Leistungen in verschiedenen Testheften zueinander in Beziehung gesetzt werden. Der Vorteil des Designs ist insbesondere darin zu sehen, dass jeder Schüler nur einen Teil der gesamten Aufgabenmenge bearbeiten muss.

Multiple-Choice-Format

Aufgabenformat, bei dem aus einer Auswahl vorgegebener Antworten eine zutreffende oder am ehesten zutreffende Antwort gewählt wird.

Nationale Bildungsstandards

Von der ständigen Konferenz der Kultusminister (KMK) für alle 16 Länder der Bundesrepublik Deutschland verbindlich verabschiedete Kompetenzerwartungen, denen die SchülerInnen bis zu einem festgelegten Zeitpunkt in ihrer Bildungsbiographie (Ende der Primarstufe, Ende der Sekundarstufe I, Ende der Sekundarstufe II) gerecht werden sollen. Diese Standards sind fachspezifisch für „Kernfächer“ festgelegt. Köller (2009) legt Wert auf die Feststellung, „...dass Bildungsstandards nicht Ursache sondern Folge von Versäumnissen im deutschen Bildungssystem sind [...]“.

Nationaler Vergleich

Schulleistungsvergleich aller 16 Bundesländer im 5 Jahres-Rhythmus durch repräsentative Stichproben der Länder.

Niedrigbegabung oder Minderbegabung

Niedrigbegabung oder Minderbegabung der Entwicklung bezeichnet eine unvollständige Entwicklung der geistigen Fähigkeiten, mit besonderer Beeinträchtigung von Fertigkeiten, die zum Intelligenzniveau beitragen (bspw. Kognition, Sprache, motorische und soziale Fähigkeiten).

Normierung

Normierung gibt auf Grundlage einer umfangreichen, repräsentativen Normierungsstichprobe Auskunft über die übliche Ausprägung des untersuchten Merkmals innerhalb einer Referenzgruppe. Die Testnormen können üblicherweise dem Testmanual bzw. den Handreichungen entnommen werden.

Normierung von Bildungsstandards

Eine der Hauptaufgaben des Institutes zur Qualitätsentwicklung im Bildungswesen (IQB) besteht darin, die von der Kultusministerkonferenz in den Jahren 2003 und 2004 verabschiedeten länderübergreifenden Bildungsstandards zu präzisieren, weiterzuentwickeln und auf deren Basis Aufgaben zu entwickeln, durch die die Erreichung der in den Bildungsstandards formulierten Kompetenzerwartungen festgestellt werden kann.

Normierungsstudie

Wissenschaftliche Untersuchung, bei dem anhand einer großen, repräsentativen Stichprobe von Probanden Normwerte für ein Testverfahren gewonnen werden.

Normorientiert

s. soziale Bezugsnorm.

Objektive Testergebnisse

Testergebnisse, die im Gegensatz zu Schulnoten unabhängig vom individuellen *bias* im Urteil der Lehrperson sind. I.d.R. handelt es sich um Ergebnisse standardisierter Schulleistungstest, bei denen soziale Vergleiche durch Angabe einer Referenzgruppe ermöglicht werden.

Objektivität

Testgütekriterium, welches erfasst, inwieweit das Testergebnis einer Person hinsichtlich der Durchführung, der Auswertung und der Interpretation von der Testsituation und vom Untersucher unabhängig ist. Wird bspw. ein standardisierter Schulleistungstest in einer Klasse von einer Lehrkraft durchgeführt, so sollte auch eine andere Lehrkraft in der gleichen Klasse die gleichen Ergebnisse erzielen.

Odds-Even Reliabilität

Die Odds-Even Reliabilität ist eine Form der Testhalbierungsreliabilität. Zur Reliabilitätsüberprüfung wird dabei der Test anhand der geraden und ungeraden Items in zwei Teile unterteilt.

Offene (Aufgaben-) Formate

Aufgabenformate, welche vergleichsweise umfangreichere, selbst formulierte Antworten erfordern und vielfältige Lösungswege ermöglichen. Dazu können Kurzaufsätze, Zeichnungen oder Formulierung von Lösungswegen gehören.

Ökonomisches Kapital

Nach Bourdieu (1983) wird ökonomisches Kapital als) Besitz jeglicher Art von Ware verstanden. Darunter fallen u. A. Produktionsmittel, Grund und Boden, und weiteres Vermögen.

Operationalisierung

Operationalisierung bezeichnet die Verknüpfung von Begriffen mit Verfahren (bzw. Operationen), so dass der Inhalt der Begriffe sich durch Empirie bestimmen lässt.

Ordinale Items

Items, welche mehrstufigen Bewertungen der Itemlösung erlauben (z.B. „Item falsch bearbeitet“ – „Item teilweise richtig bearbeitet“ – „Item richtig bearbeitet“).

Ordinalskala

Skala, bei der Differenzen zwischen den Punktwerten nicht gleiche Differenzen zwischen der Ausprägung des betrachteten Merkmals bedeuteten.

Orientierungsarbeit

s. flächendeckende Vergleichsarbeit.

Output

Die aus Lernprozessen resultierenden Ergebnisse, Kenntnisse und Fähigkeiten.

Output-Steuerung

Art der Steuerung des Bildungswesens. Vorgegeben wird, beispielsweise durch Lehrpläne, welchen Kenntnisse und Fähigkeiten die Schülerinnen und Schüler haben sollen. Gegenteil: Input-Steuerung.

Oversampling

Vergrößerung der Stichprobe in Bezug auf forschungsrelevante Merkmale. Oversampling wird häufig bei nationalen Ergänzungsstudien des Bildungsmonitorings durchgeführt, um Ergebnisse, welche sich auf das relevante Merkmal beziehen zufallskritisch abzusichern. Beispiel: Bei DESI wurden bilinguale Klassen überproportional stark in die Stichprobe aufgenommen, um sichere Aussagen über deren Leistungsfähigkeit im Vergleich zu monolingualen Klassen treffen zu können.

Pädagogische Interventionsmaßnahmen

Pädagogische Interventionsmaßnahmen sind gezielt an ein Diagnoseergebnis anschließende Förderungsaktivitäten (bspw. Vermittlung einer realistischeren Zielsetzung).

Parallelarbeit

Von allen Fachlehrerinnen und Fachlehrern einer Jahrgangsstufe in gleicher Form eingesetzte Klassenarbeit. Ihr Ziel ist die vergleichende Rückmeldung zum Lernstand unterschiedlicher Klassen.

Passwort-geschütztes Internetportal

Um Zugriff auf Begleitmaterial der Vergleichsarbeiten zu erhalten, die auf speziellen, durch die Länder eingerichteten, Internetportalen zur Verfügung stehen, kann es unter Umständen notwendig sein, diese Portale durch spezielle Zugangsdaten zu sichern und nur nach korrekter Eingabe zugänglich zu machen. Nach erfolgreicher Einwahl stehen die gewünschten Materialien zur Verfügung.

Performanz

Performanz kann als Ausdruck oder als die Anwendung von Fähigkeiten in lebensweltlich relevanten Situationen (bspw. Schule, Ausbildungssituation) beim Lernen, Wissenserwerb, in Prüfungen, der beruflichen Tätigkeit verstanden werden (vgl. Rindermann & Kwiatkowski, 2007).

Performanzziele

Performanzziele bezeichnen den Vergleich der eigenen Leistung mit den Leistungen anderer Lerner (soziale Bezugsnorm).

Pilotierung

s. Pilotierungsphase.

Pilotierungsphase

Hier: Eine Phase der Erprobung und psychometrischen Prüfung von Testaufgaben.

PIRLS

„Progress in International Reading Literacy Study“, internationale Bezeichnung für IGLU.

PISA

„Programme for International Student Assessment“. Die PISA-Studien der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) sind internationale Schulleistungsuntersuchungen, die seit dem Jahr 2000 in dreijährigem Turnus in den meisten Mitgliedstaaten der OECD und einer zunehmenden Anzahl von Partnerstaaten durchgeführt werden und zum Ziel haben, alltags- und berufsrelevante Kenntnisse und Fähigkeiten 15-Jähriger zu messen.

PISA-E

„Programme for International Student Assessment“-Ergänzungsstudie.

PISA-I-Plus

In PISA 2003 international durchgeführte Längsschnittstudie.

Plöner Beschlüsse

Beschluss der KMK, die regelmäßige Durchführung des Bildungsmonitorings mit einer Überprüfung der Bildungsstandards zu verknüpfen.

Population

Grundgesamtheit. Die Menge aller statistischen Einheiten (auch Merkmalsträger, Erhebungseinheit) mit übereinstimmenden Identifikationskriterien.

Powertest

Besitzen keine eng bemessene Zeitbegrenzung, die Komplexität der Items steigt graduell an.

Präpilotierung

Umfasst eine grobe Abschätzung der Schwierigkeit und Testgüte von Items im Vorhinein einer Pilotierungsphase.

Pre-Tests

In Pre-Tests werden die geplanten Erhebungsinstrumente verwendet, ohne dass die Ergebnisse in die Studie mit einfließen. Ziel ist es, dass für die eigentliche Untersuchung Fehlerquellen im Vorfeld erkannt und ausgeräumt werden können.

Probabilistische Testtheorie

Die probabilistische Testtheorie beschreibt den Zusammenhang zwischen Item-Antworten und nicht direkt beobachtbaren Eigenschaften wie etwa der Kompetenz von Probanden in einem bestimmten Inhaltsbereich. Eines der bekanntesten probabilistischen Modelle ist das dichotome Rasch-Modell.

Proficiency Scalings

Zuordnung von Kompetenzstufen zu Abschnitten der (kontinuierlichen) Leistungsskala

Projekt „SINUS-Transfer“

Das Projekt „SINUS-Transfer“ ist die Weiterentwicklung des Modellversuchs SINUS ("Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts"). Ziel des Projektes ist die Förderung der mathematisch-naturwissenschaftlichen Kompetenzen durch eine nachhaltige und flächendeckende Nutzung und Weiterentwicklung der Ergebnisse von SINUS.

Projekt VERA

VERgleichsArbeiten in der Grundschule sind ein von sieben Bundesländern übergreifend initiiertes Projekt in der Primarstufe, welches den Leistungsstand von Schülern in Hauptfächern (Deutsch, Mathematik) ermitteln soll. VERA wurde 2004, koordiniert von der Universität Koblenz-Landau, begonnen. Seit 2009 werden die Aufgaben für VERA vom IQB bereitgestellt.

ProLesen

ProLesen ist ein Projekt der KMK. Inhalt des Projektes ist die Förderung von Deutsch als Unterrichtssprache als Aufgabe aller Fächer.

Prozentrangnormen

Prozentrangnormen meint eine dem Rohwert entsprechende Prozentrangplatz Zuordnung. Dieser Prozentrang gibt Auskunft darüber, wie viel Prozent der Testpersonen aus der Eichstichprobe höhere oder niedrigere Werte erreicht haben. Bei Leistungstests entspricht ein hoher Prozentrang einer hohen Leistung, ein niedriger Prozentrang einer niedrigen Leistung.

Prozentrangplatz

s. Prozentrangnormen

Prozesswissen

Wissen darüber, wie man etwas tut; Wissen über Methoden des Nachforschens; Kenntnis über Anwendungskriterien für Fähigkeiten, Algorithmen, Techniken und Methoden.

Psychometriker

Ein Psychometriker befasst sich auf dem Gebiet der Psychologie mit Theorien und Methoden des psychologischen Messens.

Psychometrische Kriterien

s. Testgütekriterien.

Ranking

Kennzeichnet eine Rangordnung (auch Rangfolge, Rangliste) und ist das Ergebnis einer Sortierung von mehreren vergleichbaren Objekten, die mit einer vergleichenden Bewertung einhergeht.

Rasch-Modell

Das bekannteste Testmodell der probabilistischen Testtheorie, bei dem aus dem beobachteten Antwortverhalten zwei Parameter geschätzt werden, welche das

Antwortverhalten determinieren: die Itemschwierigkeit (Itemparameter) und die Fähigkeit (oder Kompetenz) der Person (Personenparameter). **Referenzgruppe**

Hier: Gruppe von Probanden, deren Ergebnisse eines Tests als Vergleichswerte bei Anwendung einer sozialen Bezugsnorm dienen.

Referenzwert (Fairer Vergleich)

Von fairen Vergleichen wird gesprochen, wenn in Rückmeldungen aus Leistungstests Ergebnisse für Schulklassen oder Jahrgangsstufen einem Referenzwert gegenübergestellt werden, der nicht die gesamte Population repräsentiert, sondern (a) ausschließlich Klassen oder Schulen der Stichprobe berücksichtigt, die eine ähnliche Schülerzusammensetzung haben, oder (b) die Ergebnisse anhand von leistungsrelevanten Hintergrundinformationen statistisch korrigiert.

Regelstandard

Gelten als Referenzpunkt dafür, wie empirische Ergebnisse aus Vergleichsarbeiten zu beurteilen sind, wurden in den nationalen Bildungsstandards so genannte Regelstandards formuliert, die von Schülerinnen und Schülern *durchschnittlich* erreicht werden sollen.

Reliabilität

Eines der Hauptgütekriterien eines Tests. Reliabilität entspricht dem Anteil der Streuung ins Messwerten, welcher durch tatsächliche Unterschiede in der zu messenden Eigenschaft und nicht durch Messfehler erklärt werden kann.

Retest-Reliabilität

Retest-Reliabilität kann als die zeitliche Stabilität eines Tests definiert werden. Um die Retest-Reliabilität zu erfassen wird ein Test mit der gleichen Stichprobe in einem Abstand von einigen Wochen wiederholt.

Risikogruppe

Hier: Gruppe der Schülerinnen und Schüler, denen nur wenige Kompetenzen zugeschriebene werden können und bei denen zu befürchten ist, dass sie erhebliche Schwierigkeiten haben werden, beruflichen und allgemein lebensrelevanten Anforderungen gerecht zu werden (Artelt, Stanat, Schneider & Schiefele, 2001).

Rohwert (Score)

Der Rohwert (Score) entspricht dem Testwert eines Fragebogens oder Tests.

rotierende Testhefte

s. Multi-Matrix-Design.

Rückmeldeformate

Es müssen adäquate Rückmeldeformate gewählt werden, um ein grundlegendes Verständnis der Rückmeldungen sicherzustellen

Rückmeldungen

Hier: Aufbereitet an die Schulen zurückgesandte Ergebnisse und Begleitmaterialien der Vergleichsarbeiten.

Schreiben

Hier: Eine der Dimensionen der Bildungsstandards in den Fächern Deutsch und Englisch, bei der der produktive Aspekt sprachlicher Kompetenzen Berücksichtigung findet.

Schuladministration

Unter dem Begriff der Schuladministration werden bei Kohler (2005) Beamtinnen und Beamte der Schulaufsicht verstanden.

Schulentwicklungsprozess

Als Schulentwicklung wird der systematische, zielgerichtete, selbstreflexive und für die Bildungsprozesse der Schüler funktionale Entwicklungsprozess hin zu einer Professionalisierung schulischer Prozesse bezeichnet. Schulentwicklung dient der Verbesserung der Qualität der Schule und des Unterrichts.

Schulformunterschiede

s. Leistungsstreuung.

Schulleistungsmerkmale

Schulleistungsmerkmale sind bspw. Leseleistung, Mathematikleistung oder Naturwissenschaftsleistung. Die Schulleistung ist ein Konstrukt, das im schulischen Alltag durch Zensuren operationalisiert wird. Schulleistung wird dabei unterschiedlich definiert. Die Definition verdeutlichen, dass Schulleistung durch einen dynamischen Ursachenkomplex bedingt wird, dessen multiple Faktoren von proximalen Faktoren (Denken und Wissen) bis zu distalen Faktoren (Bildungsorientierung der Eltern) reichen.

Schulleistungsrelevante Merkmale

Schulleistungsrelevante Merkmale kann als Sammelbegriff zur Erfassung sowohl kognitiver als auch motivationaler Merkmale verstanden werden. Eine präzise und eindeutige Zuordnung einzelner Merkmale in diese Kategorien ist nicht immer möglich, daher sollten kognitive und motivationale Merkmale nicht als Dichotomie, sondern eher im Sinne einer Dimension verstanden werden.

Schulübergreifende Vergleiche

Vergleich der eigenen Ergebnisse mit Ergebnissen von Referenzklassen anderer Schulen.

Schwierigkeitsbestimmende Merkmale

Item-Merkmale, welche die Lösungswahrscheinlichkeit beeinflussen.

Selbstevaluation

Evaluation, die von Personen durchgeführt wird, um die selbstverantworteten Arbeitsprozesse und -ergebnisse reflektieren und verbessern zu können.

Selbstkonzept

Selbstkonzept umfasst alle selbstbezogenen Einschätzungen, Überzeugungen und Meinungen. Es ist zu beachten, dass im wissenschaftlichen Diskurs eine Vielzahl verschiedener Modelle zum Selbstkonzept existieren (vgl. Marsh; Marsh/Shevelson).

Signifikante Abweichung

Im Zusammenhang von Vergleichsarbeiten: Abweichung einer Verteilung von SchülerInnen auf Kompetenzniveaus, welche (unter Annahme einer zuvor definierten Irrtumswahrscheinlichkeit) als statistisch bedeutsam erachtet wird.

Signifikanzniveau

Die Obergrenze für eine Irrtumswahrscheinlichkeit bei einem statistischen Testverfahren wird mit dem Signifikanzniveau (α) angegeben. Übliche Signifikanzniveaus sind 1 %, 5 % oder 10 %.

SINUS-Transfer

s. Projekt „SINUS-Transfer“.

Skala / Testskala

Satz von Items die bestimmten Bedingungen genügen müssen, um ein empirisches in ein numerischen Relativ umzuwandeln.

Skalenniveau

Statistische Eigenschaft von Merkmalen oder Variablen, welche Vorschriften für mathematisch (und logische) Operationen mit den Messwerten vorgeben. Grob wird zwischen nominalem, ordinalem und metrischem Skalenniveau unterschieden.

Skalierung

Hier: Skalierung bezeichnet die relationale Zuordnung von Merkmalsausprägungen zu einem Testwert.

Soziale Bezugsnorm

Soziale Bewertungsnormen setzen die Schülerleistung ins Verhältnis zur Leistung der gesamten Lerngruppe oder der gesamten Population.

Soziales Kapital

Soziales Kapital beschreibt nach Bourdieu (1983) die Gesamtheit aller Ressourcen, welche aus der Teilhabe an sozialen Netzwerken resultieren können.

Sozialindikatoren

Hier: Statistische Indikatoren, welche einen Rückschluss auf den sozioökonomischen Hintergrund der Schülerschaft ermöglichen.

Sozialraumstatistik

Daten der Sozialraumstatistik (Arbeitslosenquote, Quote der Sozialhilfebezieher, etc.) können unter Umständen für die Adjustierung von Testleistungen herangezogen werden.

Sozioökonomische Merkmale der Schülerschaft

Beispiele für sozioökonomische Merkmale der Schülerschaft stellen der elterliche Bildungsgrad und die Familiensprache dar. Sie beeinflussen den Aufbau von Kompetenzen und entziehen sich dem Einflussbereich von Schule und Unterricht.

Speedtest

Die Bezeichnung „Speedtest“ impliziert eine starke zeitliche Begrenzung des Tests.

Sprache (und Sprachgebrauch untersuchen)

Hier: Eine der vier Dimensionen der Bildungsstandards im Fach Deutsch stellt der Kompetenzbereich „Sprache und Sprachgebrauch untersuchen“ dar.

Standardabweichung

Die Standardabweichung ist ein sogenanntes Streuungsmaß, welches Auskunft darüber gibt, wie homogen oder heterogen eine Merkmalsverteilung ist. Je kleiner die Standardabweichung ist, desto enger gruppieren sich die Werte der einzelnen Fälle um den Mittelwert, je größer sie ist, desto weiter streuen sie um den Mittelwert.

Standards

s. Bildungsstandards / Mindeststandards / Regelstandards / Maximalstandards.

Standardüberprüfung

Summe der Maßnahmen, welche zur Überprüfung herangezogen werden können, inwiefern Kriterien eines Standards erfüllt sind.

Standortgruppe

Schulen werden Standortgruppen zugeordnet, um den Vergleich von Schulen zu ermöglichen, deren Schülerschaft möglichst große Ähnlichkeit in Bezug auf sozioökonomische Merkmale besitzt.

Standorttyp

s. Standortgruppe.

Stichprobe

In der Wissenschaft eine Teilmenge einer Population (z.B. von Schülerinnen und Schülern), welche nach bestimmten Kriterien für eine Untersuchung ausgewählt wurde. Von den Ergebnissen der Stichprobe wird – unter Berücksichtigung eines möglichen Stichprobenfehlers – auf die Population geschlossen.

Stimulus

Hier: Text, Abbildung und/oder Tabelle, welche ein Item einleiten.

Strenge oder Milde (im Urteil der Lehrperson)

Urteilsstrenge bzw.-milde der beurteilenden Lehrperson stellt bei standardisierten Tests eine (unerwünschte) Quelle von Leistungsunterschieden dar. Im Vergleich zur durchschnittlichen Beurteilung wird ein Merkmal durch die Lehrperson systematisch zu positiv (Milde) oder zu negativ (Strenge) bewertet. Daraus resultiert eine Über- oder Unterschätzung der Kompetenz der Schülerinnen und Schüler.

Streuung

Hier: Verteilung der Testergebnisse. Ein typisches Maß der Streuung ist die Standardabweichung.

Teilleistungsbereich eines Faches

s. Kompetenzbereich.

Teilleistungsstörungen

Teilleistungsstörungen werden auch als „umschriebene Entwicklungsstörungen“ bezeichnet.

Test

Ein Test ist ein wissenschaftliches Routineverfahren zur Untersuchung eines oder mehrerer empirisch abgrenzbarer Merkmale. Der Test soll eine möglichst quantitative Aussage über den relativen Grad der individuellen Merkmalsausprägung geben.

Testgütekriterien

Als Hauptgütekriterien eines Tests gelten Objektivität, Reliabilität und Validität. Als wichtige Nebenkriterien werden vor allem Testökonomie und Normierung angesehen.

Testhalbierungs-Reliabilität (Split-half Reliabilität)

Testhalbierungs-Reliabilität unterteilt den Test in zwei Hälften. Die Testhälften können als Paralleltests mit halber Länge angesehen und somit als Sonderform der Paralleltest-Methode aufgefasst werden. Für jede Testperson werden je zwei Testwerte ermittelt, die sich jeweils aus der Hälfte der Items zusammensetzen.

Testheftversionen

Mit dem Einsatz verschiedener Testhefte soll in Schulleistungsstudien der Leistungsstreuung im deutschen Schulsystem Rechnung getragen werden.

Test-Items

s. Testaufgaben.

Testökonomie

Testökonomie zählt zu den Nebengütekriterien und liegt immer dann vor, wenn ein Test einfach durchzuführen und auszuwerten ist sowie wenig Zeit und Material benötigt.

Testtheorie

Testtheorie beschäftigt sich mit der Frage des Zusammenhangs zwischen empirischen Testwerten und den zu messenden Merkmalsausprägungen (Zielkonstrukten). Gegenstand und Zielsetzung der Testtheorie ist die Bereitstellung von Gütekriterien für Testinstrumente. Unterschieden wird dabei grundsätzlich zwischen der klassischen Testtheorie und der probabilistischen Testtheorie. Als Hauptgütekriterien werden Objektivität, Reliabilität und Validität herangezogen.

TIMSS (Trends in International Mathematics and Science Study)

TIMSS ist eine internationale vergleichende Schulleistungsuntersuchung, in der Mathematik- und Naturwissenschaftsleistungen in der Grundschule, in der Sekundarstufe I und in der Sekundarstufe II untersucht werden. Sie wird seit 1995 im vierjährigen Turnus von der International Association for the Evaluation of Educational Achievement (IEA) durchgeführt.

UDiKom

Abkürzung für das Projekt „Aus- und Fortbildung der Lehrkräfte in Hinblick auf Verbesserung der Diagnosefähigkeit, Umgang mit Heterogenität, individuelle Förderung“ in dessen Rahmen die vorliegenden Studienbriefe erstellt wurden.

Unterrichtsentwicklung, datengestützt

Unterrichtsentwicklung, bei der Evaluationsergebnisse herangezogen werden, um Entwicklungsmaßnahmen zu diskutieren, z.B. wie der Kompetenzerwerb der Schülerinnen und Schülern besser gefördert werden kann.

Unterrichtssprache

Sprache, die im Unterricht für die Vermittlung von Wissen verwendet wird.

Unterstreichungsarbeiten

Itemformat, bei dem Texte, Textstellen oder Schlüsselworte markiert bzw. unterstrichen werden.

Unverfälschbarkeit

Unverfälschbarkeit liegt vor, wenn eine Testperson seine Ergebnisse nicht gezielt steuern oder verfälschen kann.

Validität

Validität eines Tests stellt das wichtigste Gütekriterium dar und sollte daher bei der Auswahl von Testinstrumenten besondere Beachtung erfahren. Die Validität gibt Auskunft über das Potential eines Testinstruments das zu messen, was er zu messen beansprucht.

Vergleichsarbeiten

Schriftliche Arbeiten der Schülerinnen und Schüler, die auf Basis vorgegebener Aufgabenstichproben landesweit in Teilleistungsbereichen ausgewählter Kernfächer (Deutsch, Mathematik und – in der Sekundarstufe - die erste Fremdsprache) mit dem Ziel durchgeführt werden, Schulleistungen orientiert an einer kriterialen und sozialen Bezugsnorm zu erfassen (vgl. Helmke & Hosenfeld, 2003).

Verteilungsorientiert

s. Normorientiert.

Zuhören

Eine der vier Dimensionen der Bildungsstandards im Fach Deutsch.

Zuordnungsaufgaben (auch Matching-Item bzw. Zuordnungs-Item)

Itemformat, bei dem die Zuordnung von Elementen verschiedener Kategorien geleistet werden muss. So können beispielsweise zwei Listen von Begriffen vorliegen, die einander zugeordnet werden müssen.

Zusatzstudien

s. Ergänzungsstudien.

Zweitsprache

Sprache, die Kinder als zweite Sprache erlernen bzw. die im sozialen Kontext außerhalb der Familie gesprochen wird.

Zyklusmodell von der Evaluation zur Innovation (nach Helmke und Hosenfeld, 2005)

Ein Modell, welches Phasen der Nutzung von Ergebnismeldungen beschreibt.