

## Beispielaufgaben Pädagogisch-psychologische Diagnostik und Evaluation (80 Minuten, 60 Punkte)

- Bei Multiple-Choice-Aufgaben (MC) werden bei 4 Alternativen jeweils 2 Punkte vergeben werden, sofern alles korrekt ist, und 1 Punkt für 3 korrekte Antworten aus 4, ansonsten 0 Punkte.
- Bei Single-Choice-Aufgaben (SC) ist jeweils 1 Punkt zu erreichen.

### Psychologische Grundlagen (der Diagnostik) und Gütekriterien

1) Das hypothetische Konstrukt "Kreativität" wird durch einen Test erfasst, in dem die Anzahl an in einer Minute notierten Verwendungsmöglichkeiten für einen Gegenstand als Indikator für die Dimension "Flüssigkeit des Denkens" dient.

a) Skizzieren Sie an diesem Beispiel die Mehrdeutigkeit von Indikatoren. (2 Punkte)

Anzahl der notierten Verwendungsmöglichkeiten könnte auch von Vorwissen, Kenntnis des Gegenstandes, Schreibgeschwindigkeit (und nicht nur von Kreativität) abhängen.

z.B. Hesse & Latzko, Kap. 2.2

b) Skizzieren Sie an diesem Beispiel den Bedeutungsüberschuss beim Konstrukt. (2 Punkte)

Kreativität ist mehr als Flüssigkeit des Denkens, somit wird niemals die ganze Kreativität erfasst.

z.B. Hesse & Latzko, Kap. 2.2

2) Selektion und Modifikation gelten als Ziele von Diagnostik. Welche Aussage(n) ist / sind korrekt? (MC)

Selektionsdiagnostik und Modifikationsdiagnostik dienen beide dazu, dass pädagogische Entscheidungen zwischen verschiedenen Handlungsalternativen nicht willkürlich, sondern begründet getroffen werden.

Modifikationsdiagnostik dient der Optimierung des Lehr- und Lernprozesses, Selektionsdiagnostik hingegen der Erteilung gesellschaftlich geforderter Qualifikationen.

Modifikationsdiagnostik ist durch sozialen Bezug und organisatorische Stabilität gekennzeichnet, Selektionsdiagnostik dagegen durch individuellen Bezug und organisatorische Flexibilität.

Selektionsdiagnostik und Modifikationsdiagnostik können beide sowohl bei Personen als auch bei Bedingungen ansetzen.

z.B. Urhahne et al., Kap. 24.4

3) Welche der folgenden Definitionen ist richtig? (SC)

- Objektivität ist die Genauigkeit oder Zuverlässigkeit, mit der das infrage stehende Merkmal gemessen werden kann.
- Reliabilität ist das Ausmaß, in dem eine Messung von den Gegebenheiten der Durchführung, der Auswertung und der Interpretation unabhängig ist.
- Testökonomie ist das Ausmaß, in dem der Test allen Probanden gleiche Chancen auf ein gutes Ergebnis ermöglicht.
- Validität ist das Ausmaß, in dem ein Verfahren tatsächlich das misst, was es zu messen vorgibt.

z.B. Urhahne et al., Kap. 24.5

4) Vor einer schriftlichen Prüfung wurde auf die Erstellung eines Notenschlüssels verzichtet. Welcher Aspekt der Objektivität kann dadurch verletzt sein? (SC)

- Auswertungsobjektivität
- curriculare Objektivität
- Durchführungsobjektivität

Interpretationsobjektivität

z.B. Urhahne et al., Kap. 25.2

5) Frau Schneider hat mit ihrer Klasse das Schreiben eines Briefes im Unterricht intensiv besprochen und prüft die Leistungen in einer Probearbeit. Da Frau Schneider möglichst genau die tatsächlichen Leistungen erfassen will, weist sie die beiden im Unterricht engagiertesten Schülerinnen gleich zu Beginn auf die formale Struktur des Briefes hin, weil sie reine Leichtsinnsfehler vermutet. Beim Korrigieren hält sich Frau Schneider an den Kriterienkatalog, der auch den Schülerinnen und Schülern vorab bekannt war. Für die Notenvergabe sortiert sie die Arbeiten nach Punktzahlen und vergibt wie immer für die drei besten Leistungen die Note 1, für die drei schwächsten die Note 5, für die restlichen 20 Arbeiten orientiert sie sich am Abstand zu den Noten 1 und 5.

a) Skizzieren Sie jeweils eine Aussage zur Durchführungsobjektivität, Auswertungsobjektivität, Interpretationsobjektivität und Validität und bewerten Sie jeweils kurz, wie die Aktivität von Frau Schneider für das Kriterium einzuschätzen ist. (8 Punkte)

Durchführung: Hilfe für 2 Schülerinnen – schlecht (2 Punkte)

Auswertung: Kriterienkatalog: gut (2 Punkte)

Interpretation: soziale Bezugsnorm – nicht vorgesehen, schlecht (2 Punkte)

Validität: Lerngelegenheit plus Transparenz der Anforderungen (gut) (2 Punkte)

- b) Für Luca gibt es beim Schreiben des Briefes einen Zeitzuschlag von 50 % aufgrund einer attestierten Lese-Rechtschreibstörung. Erörtern Sie diesen Nachteilsausgleich im Hinblick auf die Qualität der Leistungsmessung, indem Sie auf zwei Gütekriterien eingehen. (4 Punkte)

Gut für Validität, da nicht Schreibgeschwindigkeit erfasst werden soll – kritischere Aussagen wären möglich (2 Punkte)

Mögliches Problem für Durchführungsobjektivität, da unterschiedliche Bedingungen – ließe sich aber auch so darstellen, dass Bedingungen für Leistungserbringung (unabhängig von Schreibgeschwindigkeit) dadurch sogar angenähert werden (2 Punkte)

- c) Luca erhält trotz des Zeitzuschlages nur die Note 4, was für Luca eine große Enttäuschung darstellt. Skizzieren Sie im Hinblick auf Bezugsnormorientierung und Kausalattribution / Ursachenerklärung jeweils eine Möglichkeit, wie Frau Schneider die Leistung motivationsförderlich kommentieren könnte. (4 Punkte)

Verwendung individueller Bezugsnorm (Ansätze zur Verbesserung erkannt; 2 Punkte) – Misserfolg auf variable, kontrollierbare Faktoren zurückführen (nicht gut vorbereitet, ungünstige Strategie, ...; 2 Punkte).

mehrfach; Aufgabe im offiziellen Schreiben an die Studierenden

### Prinzipien der Standardisierung und Normierung

- 5) Was trifft auf Normwerte standardisierter psychologischer Testverfahren zu? (MC)

- Normwerte ermöglichen den Vergleich von Leistungen mit einer repräsentativen Vergleichsstichprobe.
- Normwerte stellen eine Wahrscheinlichkeitsaussage über die Häufigkeit bzw. Seltenheit eines bestimmten Leistungsergebnisses dar.
- Ein Normwert ist ein Schätzwert für die alters- oder klassenstufenspezifische Ausprägung einer latenten Eigenschaft.
- Allen Normwerten liegt eine kriteriale (sachliche) Bezugsnorm zugrunde.

z.B. Hesse & Latzko, Kap. 2.2; Inenkamp & Lissmann, Kap. 2.4

## Bezugsnormen

6) Welche Bezugsnormorientierung wäre für einen Schüler, der im Englischunterricht die Symptome der erlernten Hilflosigkeit zeigt, wahrscheinlich am förderlichsten? (SC)

- die individuelle Bezugsnorm
- die sachliche Bezugsnorm
- die soziale Bezugsnorm
- die psychologische Bezugsnorm

mehrfach; Aufgabe im offiziellen Schreiben an die Studierenden

## Methoden der Diagnostik

7) Skizzieren Sie jeweils einen Grund, warum ein Schulleistungstest eine höhere, aber auch eine niedrigere Validität als eine schulische Probearbeit aufweisen könnte. (4 Punkte)

höher, weil hoffentlich aufwändig mit Expertinnen und Experten konstruiert auf Basis der Lehrpläne

niedriger, weil möglicherweise fehlende Lerngelegenheit / Probearbeit am konkreten Unterricht orientiert

z.B. Hesse & Latzko, Kap. 3.2

8) Bei der Entscheidung, ob ein Kind auf ein Gymnasium gehen sollte, ... (MC)

- kann Selektionsdiagnostik (z.B. Bedingungsselektion) eine Rolle spielen.
- dürfen auch die Ergebnisse standardisierter Testverfahren in Betracht gezogen werden.
- dürfen nur kognitive Variablen den Ausschlag geben, weshalb auf die Erfassung nicht-kognitiver Variablen objektiverweise verzichtet werden muss.
- sollten Intelligenztests eingesetzt werden und die bisherigen Leistungen („Vorläuferfertigkeiten“ bzw. Schulnoten) weitgehend ausgeblendet werden.

z.B. Ingenkamp & Lissmann, Kap. 4.3

9) Skizzieren Sie drei Nachteile, wenn ausschließlich standardisierte pädagogisch-psychologische Testverfahren zur Vergabe von Noten verwendet werden würden. (6 Punkte)

u.a. Tests für viele Bereiche gar nicht vorhanden – mangelhafte Validität, da teils fehlende Lerngelegenheit – Unterricht könnte erstarren, wenn nur auf Tests ausgerichtet

z.B. Hesse & Latzko, Kap. 2.2

## Diagnostische Kompetenz von Lehrkräften

10) Obwohl Klaus und Klara in der Biologie-Aufgabe zur Beschreibung des Mitochondriums inhaltlich die gleichen Aspekte beschrieben haben, erhält Klara eine bessere Note als Klaus und zusätzlich den Hinweis, ein besonders schön zu lesendes Schriftbild zu haben.

a) Welcher Beurteilungsfehler wurde hier vermutlich von der Lehrkraft begangen? (SC)

Erinnerungsfehler

Reihenfolgefehler

Kontrasteffekte

Halo-Effekt

z.B. Hesse & Latzko, Kap. 1.5

b) Stellen Sie einen empirischen Befund dar, der den Einfluss von Beurteilungsfehlern bei der Bewertung schulischer Leistungen aufzeigt! (4 Punkte)

z.B. Einfluss der Sprechgeschwindigkeit auf Note in mündlicher Prüfung (Birkel & Pritz, 1980) (1 Punkt für Autorennamen, 1 Punkt für Design, 1 Punkt für zentralen Befund, 1 Punkt für Details, z.B. ungefähre Effektstärke / Unterschied in Notenmittelwert ...)

z.B. Ingenkamp & Lissmann, Kap. 3.2

c) Skizzieren Sie drei Maßnahmen, die die Lehrkraft ergreifen könnte, zukünftig vergleichbare Urteilsfehler zu vermeiden. (3 Punkte)

Standardisierung der Auswertung durch Kriterienkatalog

vorab festgelegter Notenschlüssel

Korrektur Aufgabe für Aufgabe mit Mischung der Reihenfolge

Verwendung MC

mehrfach

## Diagnostik von Bedingungen des Lernens

11 a) Nennen Sie einen Intelligenztest und den Altersbereich, für den er normiert ist. (2 Punkte)

z.B. KFT 4-12+, 4.-12./13. Klasse

z.B. Hesse & Latzko, Kap. 3.1.1.1

b) Was bedeutet es, wenn eine Schülerin in diesem Test einen IQ von 94 erreicht hat? (3 Punkte)

Sie ist im Durchschnittsbereich, besser als etwa 35 % [25-45%] der Vergleichsgruppe. (1 Punkt für Vergleichsgruppe oder Ähnliches, 1 Punkt für Durchschnittsbereich, 1 Punkt für konkretere Angabe)

z.B. Ingenkamp & Lissmann, Kap. 2.4

12 a) Skizzieren Sie eine Herangehensweise, um das fachspezifische Selbstkonzept eines Schülers mit unterdurchschnittlichen Schulleistungen zu erfassen. (4 Punkte)

z.B. Einsatz eines standardisierten Fragebogens, Selbsteinschätzung anhand mehrerer Items, bei SESSKO in Bezug auf verschiedene Bezugsnormen

z.B. Hesse & Latzko, Kap. 3.1.1.3; bei den Zusatzinfos

b) Wie schätzen Sie die Validität dieser Herangehensweise ein? (4 Punkte)

Bei standardisiertem Fragebogen (SESSKO) gute Grundlage durch Objektivität und Reliabilität, tatsächlich das Selbstkonzept erfassende Items, Gefahr sozialer Erwünschtheit

z.B. Hesse & Latzko, Kap. 3.1.1.3; bei den Zusatzinfos

c) Skizzieren Sie zwei zentrale Maßnahmen, um ein sehr negatives Selbstkonzept positiv zu beeinflussen. (6 Punkte)

z.B. individuelle Bezugsnormorientierung bei Leistungsrückmeldungen, also Vergleich mit bisherigen Leistungen, wodurch es mittelschwer ist, Erfolg zu haben

z.B. Rückmeldungen, die günstige Kausalattribution nahelegen, kontrollierbare Faktoren wie ungünstige / mangelnde Vorbereitung bei Misserfolg – gute Vorbereitung / Kompetenzzuwachs bei Erfolg – da Selbstkonzept gefragt auch ok: external bei Misserfolg, internal bei Erfolg

z.B. Hesse & Latzko, Kap. 3.1.1.3

### **Diagnostik von Lernprozessen; Diagnostik von Lernergebnissen und Schulleistungen**

13) Skizzieren Sie jeweils eine spezifische Maßnahme, um bei einer mündlichen Prüfung eine möglichst hohe a) Durchführungsobjektivität, b) Auswertungsobjektivität und c) Validität zu erreichen. (6 Punkte)

a) standardisierte Durchführung, z.B. zufälliges Ziehen vorbereiteter Fragen

b) Standardisierung der Auswertung, z.B. Kriterienkatalog

c) vorab ausgewählte Fragen an Prüfungsgrundlage / Unterricht orientiert – mündlich dann prüfen, wenn Sprache Gegenstand der Prüfung

z.B. Urhahne et al., Kap. 24.5; bei den Zusatzinfos

14) Emma erreicht in einem Schulleistungstest für Mathematik einen Prozentrang (PR) von 75. Ihre Mathematiknoten liegen in diesem Schuljahr zwischen 3 und 4. Welche Aussage(n) trifft / treffen zu? (MC)

Die Bestimmung des Prozentrangs muss im Rahmen der qualitativen Auswertung des Tests erfolgt sein.

Emma hat im Test besser (oder genauso gut) abgeschnitten als 75 % im Rahmen der Eichstichprobe getesteten Schülerinnen und Schüler ihres Jahrgangs.

Möglicherweise erfolgt die Notenvergabe recht streng.

Im Gegensatz zu den Noten hat ein Schulleistungstest keinen Lehrplanbezug.

z.B. Hesse & Latzko, Kap. 2.2; Aufgabe im offiziellen Schreiben an die Studierenden

15) Welche Aussage(n) zur Verteilung von Punktzahlen in schulischen Probearbeiten ist / sind zutreffend? (MC)

Punktzahlen in schulischen Probearbeiten müssen normalverteilt sein, damit Noten gegeben werden können.

Durch eine Häufung besonders starker und besonders schwacher Leistungen entsteht eine Normalverteilung.

Sind viele Leistungen fast perfekt, während einzelne Arbeiten deutliche Defizite aufweisen, deutet das auf eine Abweichung von einer Normalverteilung hin.

Wenn viele voneinander unabhängige Faktoren zur Leistung beitragen, ist eine Normalverteilung extrem unwahrscheinlich.

z.B. Hesse & Latzko, Kap. 2.2

16) Nennen Sie einen Grund, warum ein Mittelwert aus zwei fünfminütigen Abfragen eine bessere Reliabilität als eine einzelne fünfminütige Abfrage haben dürfte. (2 Punkte)

Messfehler können sich ausgleichen; länger ist besser für die Reliabilität.

z.B. Urhahne et al., Kap. 25.2; bei den Zusatzinfos

17) Ist bei der Benotung von Schulaufgaben die soziale oder die kriteriale Bezugsnorm vorzuziehen? Begründen Sie kurz Ihre Entscheidung (2 Argumente)! (4 Punkte)

Kriteriale, da bereits so gesetzlich vorgegeben – Schulklasse keine geeignete Stichprobe für soziale Bezugsnorm

## Diagnostik von Lern- und Leistungsschwierigkeiten

18) Jona hat Schwierigkeiten im Lesen und Rechtschreiben. Eine Lese-Rechtschreibstörung wurde nicht abgeklärt.

a) Skizzieren Sie, welche Information das Ergebnis (z.B. der Prozentrang) eines normierten Lese- bzw. Rechtschreibtests liefern könnte, die Sie nicht in schulischen Leistungserhebungen bekommen könnten. (3 Punkte)

Überregionaler Vergleich müsste dabei sein, also Vergleich mit Normstichprobe

b) Die Leistungen von Jona liegen im Grenzbereich zwischen noch durchschnittlichen und unterdurchschnittlichen Leistungen. Zu welchem Prozentrang würde diese Aussage am besten passen? (SC)

PR 1

PR 4

PR 16

PR 32

PR 48

PR 88

PR 92

c) Beschreiben Sie weitere Informationen, die Sie durch eine differenziertere qualitative Auswertung des Tests erhalten könnten. (5 Punkte)

Hinweise auf verwendete Strategie beim Lesen, evtl. Lesegenauigkeit, Lesegeschwindigkeit, Leseverständnis, ..., Fehlerschwerpunkte beim Rechtschreiben, z.B. Verstöße gegen die Lauttreue, bestimmte Regeln, ...

d) Jonas Lehrkraft verzichtet auf die Bewertung von Rechtschreibfehlern in Heimat- und Sachkunde, obwohl kein Attest auf Lese-Rechtschreibstörung vorliegt. Erläutern Sie jeweils eine mögliche Auswirkung auf die Auswertungsobjektivität und auf die Validität. (4 Punkte)

Eigentlich keine Auswirkung auf Auswertungsobjektivität, vor allem, wenn Lehrkraft generell Rechtschreibung nicht bewertet. (2 Punkte)

Gut für die Validität, da nur HSK-Leistung erfasst werden soll. (2 Punkte)

z.B. Hesse & Latzko, Kap. 3.2 / 3.4; bei den Zusatzinfos

19) Erläutern Sie, welche diagnostischen Verfahrensweisen sich für Sie als Lehrkraft anbieten, wenn Sie bei einer Schülerin / einem Schüler mit Lernschwierigkeiten genauere Informationen über die Problematik sowie mögliche Interventionsmöglichkeiten gewinnen möchten! (12 Punkte)

max. 4 Punkte für Verfahren zur genaueren Analyse der Schwierigkeiten (Tests (Schulleistungstests) mit qualitativer Analyse / z.B. Fehleranalyse beim Rechtschreiben; Analyse der Fehler in schulischen Leistungserhebungen; Erfassen von Teilfähigkeiten, z.B. beim Rechnen)

max. 8 Punkte für mindestens 4 Faktoren, die an Lernschwierigkeiten beteiligt sein könnten, jeweils inklusive passendem Verfahren (Ängstlichkeit; Selbstkonzept; z.B. jeweils Fragebogen; Lernstrategien, z.B. Fragebogen, Beobachtung; Kausalattributionsmuster, z.B. Interview u.a.; auch häusliches Umfeld, Unterstützung, z.B. Elterngespräch)

z.B. Hesse & Latzko, Kap. 3.2 / 3.4

### **Methoden der schulbezogenen Evaluation und Bildungsmonitoring; Evaluation von Lernumgebungen und Unterricht mit digitalen und analogen Medien**

20) Evaluation auf der Ebene des Bildungssystems. In den PISA-Studien sollen Kompetenzen erfasst werden ("Literacy"-Ansatz). Welche Aussage charakterisiert diesen Ansatz am besten? (SC)

Erfasst werden soll die allgemeine intellektuelle Leistungsfähigkeit.

Es sollen diejenigen Leistungsbereiche überprüft werden, die im Lehrplan vorgegeben sind.

Der Ansatz zielt darauf, das zu erfassen, was im Unterricht typischerweise vermittelt wird.

Es sollen Kompetenzen im Sinne einer Grundbildung erfasst werden, insbesondere die Fähigkeit, Wissen auch anzuwenden.

z.B. Urhahne et al., Kap. 26.3

21) Wie unterscheiden sich formative und summative Evaluation? (MC)

Eine formative Evaluation betrachtet eher den Prozess, während eine summative Evaluation eher das Ergebnis in den Blick nimmt.

Eine formative Evaluation basiert in der Regel auf Testverfahren, die summative Evaluation dagegen auf Verhaltensbeobachtung.

Eine formative Evaluation richtet sich eher an Durchführende während des Prozesses, während die summative Evaluation eher im Sinne einer abschließenden Bewertung an Auftraggeber / Entscheidungsebene berichtet wird.

Formativ wird eher extern evaluiert, während interne Evaluationen summativ sind.

z.B. Urhahne et al., Kap. 26.2

24) Nennen Sie zwei Evaluationsstandards und erläutern Sie einen davon kurz. (4 Punkte)

Nützlichkeit / Durchführbarkeit / Fairness / Genauigkeit (2 Punkte fürs Nennen)

z.B. Nützlichkeit durch Ausrichtung am Informationsbedarf der Stakeholder

z.B. Urhahne et al., Kap. 26.2

25) Beurteilung der Unterrichtsqualität – Erläutern Sie zwei Vor- und zwei Nachteile der Unterrichtsbeobachtung (durch eine andere Lehrkraft) gegenüber Schüler:innenfeedback. (12 Punkte)

Nachteile (= Vorteile Schüler:innenfeedback): SuS direkt am Unterrichtsgeschehen beteiligt / ganzes Schuljahr Basis der Bewertung möglich / Zusammenfassung unterschiedlicher Einzelperspektiven / Aufwand (max. 6 Punkte, je 2 pro passendes Argument)

Vorteile: Didaktische Kompetenz und fachliche Expertise durch andere Lehrkraft eher beurteilbar / systematische Beobachtung reduziert Bewertungsverzerrungen / auch Konfundierung mit Beliebtheit bei Schüler:innenfeedback u.a. (max. 6 Punkte, je 2 pro passendes Argument)

z.B. Hesse & Latzko, Kap. 3.1.2; bei den Zusatzinfos