



Bild von [F1 Digitals](#) auf [Pixabay](#)

# Was Sie schon immer wissen wollten über Entwicklungs-, Intelligenz- und Schultests einschließlich pädagogischer Beobachtungsbögen

von *Frank Herold*

## **1. Einleitung**

Was genau sind Entwicklungs- und Intelligenztests, Schultests und Beobachtungsbögen? Und was sind ihre fachlichen Hintergründe? Wenn Sie das interessiert und Sie einfache Darstellungen bevorzugen, dann sind Sie hier richtig.

Sie erfahren z.B. etwas darüber, inwieweit man Merkmale wie Intelligenz überhaupt messen kann; Sie lernen kennen Bewertungsmaßstäbe oder Skalen, Grundgesamtheit und Stichprobe, Standard- und Prozentrangnormen sowie die Ungenauigkeit von Tests und Beobachtungsbögen einschließlich ihrer wichtigsten Qualitätsmerkmale. Gerechnet wird kaum, und wenn, dann nur mit Grundrechenarten. Zum guten Schluss geht es darum, wie man Kinder auf einen Test vorbereitet, ob ein Elternteil beim Test anwesend sein sollte und wie man das Testergebnis vermittelt.

## **2. Psychologischer Test – was ist das? Erste Hinweise.**

Ein psychologischer Test ist ein Prüfverfahren für praktische oder wissenschaftliche Untersuchungszwecke. Mit ihm soll man Personen danach unterscheiden können, wie

---

gew-ansbach.de – 29.10.2023 – S. 1

Was Sie schon immer wissen wollten über Entwicklungs-, Intelligenz- und Schultests einschließlich  
pädagogischer Beobachtungsbögen

---

ausgeprägt bestimmte Merkmale bei ihnen sind. Das ist vergleichbar mit einer Klassenarbeit, mit der man Wissensunterschiede zwischen Schülerinnen und Schülern erkennen kann (siehe Fisseni 1997, 25).

Auch unsere Tests vergleichen die Leistungen oder Verhaltensweisen eines Kindes oder jugendlichen Menschen mit anderen Kindern und Jugendlichen, also mit einer „Vergleichsgruppe“ (Eikenbusch/Leuders 2008, 206).

Hier unterscheiden sich aber auch Klassenarbeit und psychologischer Test: Bei einer Klassenarbeit ist die Vergleichsgruppe die Klasse. Dagegen soll ein Test für viel größere Gruppen gelten, z.B. für alle vier- bis sechsjährigen Kinder in Deutschland. Die müsste man alle untersuchen, um vollzählige Vergleichswerte zu haben. Das geht aber nicht. Also untersucht man nur eine ausgewählte Teilgruppe dieser Kinder, die sogenannte Normstichprobe (Eichstichprobe) - siehe Abschnitt 7. Mit ihr kann man dann die Leistung jedes vier- bis sechsjährigen Kindes in Deutschland einschätzen als unterdurchschnittlich, durchschnittlich oder überdurchschnittlich.

Außerdem gilt für psychologische Tests, dass die Bedingungen bei Durchführung und Auswertung standardisiert sind, also vereinheitlicht - siehe Abschnitt 10.1. Und schließlich drückt man das Testergebnis meist in Zahlen aus (siehe Perleth o. J., 2).

Im weiteren Sinne zählen zu psychologischen Tests z.B. auch standardisierte oder vereinheitlichte Beobachtungen (siehe Bühner 2011, 31).

## **2.1. Entwicklungstests**

Mit Entwicklungstests kann man prüfen, ob sich ein Kind altersgemäß entwickelt. Es gibt allgemeine und spezielle Entwicklungstests.

**Allgemeine Entwicklungstests** sollen möglichst alle wesentlichen Bereiche der kindlichen Entwicklung erfassen (siehe Koglin/Hallmann/Petermann 2011, 6). Das betrifft z.B. die motorische, sprachliche und soziale Entwicklung.

Die Leistung eines Kindes wird verglichen mit den Leistungen gleichaltriger Kinder. Das geschieht über die altersbezogenen Vergleichs-Normen aus einer Normstichprobe, vor allem über EQ-Werte. EQ steht für Entwicklungsquotient. Neben einem Gesamt-EQ gibt es ein Entwicklungsprofil. Es zeigt die einzelnen Entwicklungsbereiche so, dass man Stärken und Schwächen erkennen kann (siehe Macha 2012, Entwicklungsprofil). Das kann dazu beitragen, Entwicklungsprobleme abzuklären und Förderhinweise zu geben.

Beispiel: Petermann, F., Macha, T. (2013): Entwicklungstest für Kinder von sechs Monaten bis sechs Jahren - Revision (ET 6-6-R). Frankfurt a. M.: Pearson, 2., korrigierte Auflage 2015.

**Spezielle Entwicklungstests** sind für einzelne Bereiche wie Motorik oder Sprache.

## **2.2. Entwicklungs-Screenings (Kurztests)**

Entwicklungs-Screenings sind dafür da, mit wenigen Aufgaben frühzeitig diejenigen Kinder zu erkennen, die in ihrer altersgemäßen Entwicklung bedroht sind. Dabei untersucht man natürlich nur Merkmale, die entscheidend sind für die künftige Entwicklung.

Im Allgemeinen geht es lediglich um eine Unterscheidung von zwei Ergebnissen, z.B. „auffällig“ oder „unauffällig“ (siehe Macha 2007a). Ein auffälliges Ergebnis wird durch eine ausführliche Diagnostik bestätigt oder widerlegt (siehe Flender 2005, 14). Weniger gut einschät-

zen kann man mit Screenings die Fortschritte eines gesunden Kindes oder die Entwicklung eines Kindes mit Behinderung (siehe Pauen 2017, 174).

Beispiel: Tröster, H., Flender, J., Reineke, D., Wolf, S. M. (2016): DESK 3-6 R.

Dortmunder Entwicklungsscreening für den Kindergarten – Revision. Göttingen: Hogrefe.

### 2.3. Intelligenztests

Mit Intelligenztests kann man geistige Fähigkeiten, Begabungen und Schwächen erkennen (siehe Petermann/Daseking 2015, 119).

Gestatten Sie mir hier eine persönliche Frage: Möchten Sie gerne wissen, wie intelligent Sie sind? Oder doch lieber nicht - vorsichtshalber? Keine Sorge: Wie es um Ihre Intelligenz insgesamt bestellt ist, kann niemand herausfinden.

Sogar die Entwicklerinnen und Entwickler von Intelligenztests gehen davon aus, dass ihre Tests nicht „die“ Intelligenz erfassen, sondern nur einen Teilbereich der Intelligenz (siehe Joel 2018, 198). Warum ist das so?

„Es gibt viele Auffassungen darüber, was man unter Intelligenz versteht“ (Schmidt-Atzert/Krumm/Amelang 2021, 254). Intelligenz ist auch kultur- und situationsabhängig und beinhaltet im australischen Busch sicher andere Fähigkeiten als im Büro einer Bank (siehe Eikenbusch/Leuders 2008, 45). Intelligenz kann man eben einfach nicht messen in all ihrer Vielschichtigkeit (siehe Nguyen-Kim 2021, 205).

Dagegen kann man Teilbereiche der Intelligenz vergleichsweise gut einschätzen.

Unterschiedliche Intelligenztests erfassen - je nach Aufgabenart - unterschiedliche Teilbereiche (siehe Zimmermann 2011, 14). So kann eine Person in zwei unterschiedlichen Tests einmal einen hohen und einmal einen deutlich niedrigeren Intelligenzwert (IQ) erzielen (siehe Schmidt-Atzert/Amelang 2012, 42).

**Breit angelegte Intelligenztests** erfassen verschiedene - aber eben nie alle - Bereiche der Intelligenz, z.B. Gedächtnis, räumliches und schlussfolgerndes Denken, sprachliche und rechnerische Fähigkeiten.

Der Intelligenzquotient oder IQ gibt an, wo eine Person steht mit ihrer geistigen Leistungsfähigkeit, bezogen auf die altersgleiche Normstichprobe des durchgeführten Tests (siehe Zimmermann 2011, 3). Es gibt einen Gesamt-IQ und ein Intelligenzprofil.

Beispiel: Wechsler, D. (2017). Wechsler Intelligence Scale for Children – Fifth Edition (WISC-V). Deutsche Bearbeitung durch F. Petermann. Frankfurt a. M.: Pearson.

**Schmal angelegte Intelligenztests** erfassen einen kleinen Ausschnitt der Intelligenz.

Beispiel Grundintelligenz: Diese grundlegende Denkfähigkeit ist vergleichsweise frei von Sprache, Kultur und Bildung des Kindes (siehe Macha 2009, Grundintelligenztest Skala 1). Man erfasst sie mit Testaufgaben zum schlussfolgernden und räumlichen Denken mit Figuren, ohne sprachliche Aufgaben, ohne Rechenaufgaben (siehe Perleth o. J., 6). Solche Tests sind auch geeignet für Kinder mit nichtdeutscher Muttersprache und für Kinder, die aus anderen Gründen sprachliche Probleme haben.

Beispiel: Weiß, R. H., Osterland, J. (2013). CFT 1-R. Grundintelligenztest Skala 1 - Revision. Göttingen: Hogrefe.

## 2.4. Schultests

Schultests erfassen die schulischen Kenntnisse und Fähigkeiten einzelner Schülerinnen und Schüler, z.B. im Lesen, Rechtschreiben und Rechnen.

Lehrpersonen beurteilen die Leistungen ihrer Schützlinge zumeist durch Vergleiche innerhalb der Schulklasse; dagegen verwenden Schultests in der Regel Vergleichs-Normen einer Normstichprobe, die alle Bundesländer vertreten soll (siehe Schmidt-Atzert/Krumm/Amelang 2021, 675). Die Normen sind nicht altersbezogen, sondern beziehen sich auf ein Schulfach, einen Schultyp und eine Jahrgangsstufe (siehe Trollenier o. J., Abschnitt 2.1.3.). Sie werden in Form von T-Werten oder Prozenträngen (PR) wiedergegeben, siehe die Abschnitte 8.1. und 8.2.

Mit Schultests klärt man häufig „Teilleistungsschwächen“ ab: Die Kinder haben besondere Probleme beim Erlernen des Lesens, Rechtschreibens oder Rechnens, obwohl sie angemessen beschult werden, intelligent genug sind und kerngesund. Stichworte: Lese-Rechtschreib-Schwäche/Störung und Rechen-Schwäche/Störung.

Beispiel: Roick, T., Göllitz, D., Hasselhorn, M. (2018): DEMAT 3+. Deutscher Mathematiktest für dritte Klassen. Göttingen: Hogrefe, 2., überarbeitete Auflage mit erweiterten und überprüften Normen.

## 2.5. Beobachtungsbögen zur Einschätzung der kindlichen Entwicklung

Mit den hier angesprochenen Beobachtungsbögen kann man die Entwicklung von Kindern in Kindertageseinrichtungen einschätzen.

Sie sind eine Zusammenstellung altersentsprechender Entwicklungsaufgaben und Verhaltensweisen, je nach Schwerpunkt z.B. zu Motorik, Sprache, Sozialverhalten oder Schulfähigkeit. Die Fachleute erfassen dann, wie gut ein Kind die Aufgaben löst oder wie häufig es diese Verhaltensweisen zeigt (siehe Flender/Tröster 2005, 10). Viele Bögen umfassen mehrere Bereiche.

Die Altersangaben beruhen im besten Falle auf Altersnormen einer Normstichprobe. Sie zeigen, welche Leistungen Kinder der jeweiligen Altersgruppe im Allgemeinen erbringen. Dadurch kann man das Ergebnis eines Kindes vergleichen mit dem Ergebnis Gleichaltriger (siehe Flender/Tröster 2005, 11). So kann man die Kinder gezielter unterstützen und die Eltern besser beraten. Hinweise auf Probleme erfordern eine genaue Diagnostik.

Beispiel: Mayr, T., Bauer, C., Krause, M. (2014): KOMPIK - Kompetenzen und Interessen von Kindern. Beobachtungs- und Einschätzungsbogen für Kinder von 3,5 bis 6 Jahren. Begleitendes Handbuch für pädagogische Fachkräfte. Gütersloh: Bertelsmann Stiftung. (KOMPIK ist kostenlos verfügbar unter: [www.kompik.de](http://www.kompik.de))

So, das war ganz schön viel für den Anfang, oder? Eine gute Gelegenheit für eine Pause. Ich habe mir Pausen meist veredelt mit etwas Schokolade, was wohl auf den nächsten Abschnitt abgefärbt hat.

## 3. Merkmale und Merkmalsausprägungen

Vollmilkschokolade ist süß, das ist eines ihrer geschätzten Merkmale. Der Süßgeschmack kann unterschiedlich ausgeprägt sein, je nach Hersteller. Leider spielen Schoko-Merkmale hier keine weitere Rolle. Doch auch Merkmale wie soziale Fähigkeiten, Intelligenz und Leseverständnis haben ihren Reiz und können sehr unterschiedlich ausgeprägt sein.

#### **4. Bewertungsmaßstab, Skala**

Um zu messen, wie ausgeprägt ein Merkmal ist, braucht man einen Bewertungsmaßstab, eine Skala. Eine Länge messen wir z.B. mit der Zentimeter-Skala eines Bandmaßes, und die Temperatur lesen wir ab von der Celsius-Skala eines Thermometers. Länge und Temperatur sind Beispiele für physikalische Merkmale, die man durch Messungen direkt bestimmen kann (siehe Dinter 2011, 1).

Menschliche Fähigkeiten - z.B. soziale Fähigkeiten, Intelligenz oder Leseverständnis - lassen sich aber nicht so direkt und zuverlässig messen wie viele physikalische Merkmale. Sie lassen sich auch nicht ablesen an einem Messgerät. Stattdessen muss man Menschen genau beobachten in ihrem Verhalten oder sie befragen (siehe Eikenbusch/Leuders 2008, 64). Für die Ergebnisse gibt es dann auch Skalen.

Beispiel IQ-Skala: Intelligenztests haben unterschiedlich schwere Aufgaben, die gelöst werden müssen (siehe Zimmermann 2011, 18). Die Leistung, die eine Person dabei zeigt, ist ihre Intelligenzleistung. Diese Leistung rechnet man um in IQ-Werte, und zwar anhand der Leistungen einer gleichaltrigen Normstichprobe. „Intelligenz lässt sich also durch IQ-Tests indirekt beobachten“ (Nguyen-Kim 2021, 212).

(Angesprochen ist hier die „probabilistische Testtheorie“. Sie enthält Theorien, die den Zusammenhang beschreiben zwischen Testergebnissen - gewonnen aus beobachtbarem Testverhalten - und den getesteten Merkmalen wie Intelligenz, die eben nicht direkt beobachtbar sind; siehe z.B. Jürgens/Lissmann 2015, 136 ff.).

Testskalen wie die IQ-Skala oder EQ-Skala zählt man zu den Intervallskalen - siehe Abschnitt 6. Vorher erzähle ich Ihnen aber etwas über die einfacheren Rangskalen, die Ihnen in der einen oder anderen Form schon vertraut sind.

#### **5. Was ist eine Rang- oder Ordinalskala?**

Beispiel Platzierung im Wettkampf, Marathonlauf: Platz 1 ist für die Läuferin oder den Läufer mit der kürzesten Zeit bis zum Ziel, Platz 2 ist für die nächste Läuferin oder den nächsten Läufer usw. Das ergibt eine sinnvolle Rangreihe: Platz 1, Platz 2, Platz 3 usw. Wichtig: Die Abstände zwischen den Zahlen - also zwischen 1, 2 und 3 usw. - sagen rein gar nichts aus über die Merkmalsunterschiede in Form der gelaufenen Zeiten (siehe Freie Universität Berlin o. J., 16, Ordinalskala). So können z.B. vier Minuten liegen zwischen den Plätzen 1 und 2 und fünfzehn Minuten zwischen den Plätzen 2 und 3.

Allgemein gesprochen erlauben die Zahlen einer Rangskala nur Aussagen im Sinne von „größer oder kleiner“ und „nach oder vor“ (siehe Bartel 1971, 12, Abschnitt 1.3.3.).

##### **5.1. Was ist eine Rang- und Schätzskala? Darf man mit ihren Werten rechnen?**

Das erklärt sich leicht am Beispiel der Schulnoten-Skala. Sie ist eine Rangskala (siehe Wirtz/Nachtigall 2008, 51). Und sie ist eine schulische Schätzskala.

Ihre Ziffern 1, 2, 3, 4, 5 und 6 sind nur die Abkürzungen für die sprachlichen Noten sehr gut, gut, befriedigend, ausreichend, mangelhaft und ungenügend (siehe Eikenbusch/Leuders 2008, 26).

Auch Schulnoten bilden eine sinnvolle Rangreihe. Schon Grundschulkindern wissen: Eine 1 ist besser als eine 2, eine 2 ist besser als eine 3 usw. Was sie noch nicht wissen: Die Abstände

zwischen den Schulnoten sind überhaupt nicht festgelegt. So bleibt z.B. völlig offen, ob der Leistungsunterschied zwischen den Noten 2 und 3 genauso groß ist wie zwischen den Noten 4 und 5, oder ob er kleiner oder größer ist (siehe Zimmermann 2010, 1). Schulnoten sind eben „keine Messwerte, sondern Schätzwerte“ (Zimmermann 2011, 10).

Genau genommen darf man deshalb nicht rechnen mit Schulnoten und auch keine Durchschnittsnoten berechnen (ebd.). Denn eine Durchschnittsnote ist eine Art von Mittelwert, ein **arithmetisches Mittel**: Man zählt alle Werte zusammen und teilt das Ergebnis durch die Anzahl der Werte.

Beispiel: Ein Kind bekommt in fünf Aufsätzen die Noten 2, 2, 3, 3, 5. Das arithmetische Mittel - die Durchschnittsnote - ist:  $(2+2+3+3+5)$  geteilt durch  $5 = 3$ .

Wie gesagt: Mathematisch erlaubt ist das nicht bei Rangskalen. Deshalb kann das arithmetische Mittel auch nur ein vager Anhaltspunkt sein für die Einschätzung von Schulleistungen (siehe Eikenbusch/Leuders 2008, 26).

Gibt es denn auch mathematisch korrekte Mittelwerte bei Rangskalen? Gewiss! Sogar zwei. Da ist zunächst der Modalwert als häufigster Wert einer Häufigkeitsverteilung. Uns interessiert aber nur der Median, der in Abschnitt 8.2. zur Geltung kommt.

## 5.2. Schätzskaalen von Beobachtungsbögen

Beispiel soziale Entwicklung: Das Kind hält sich an die Gruppenregeln.

1 = nie    2 = selten    3 = manchmal    4 = oft    5 = sehr oft

Schätzskaalen sind in den meisten Fällen nur Rangskalen (siehe Bühner 2011, 114). Denn die Abstände ihrer Antwortstufen kann man meist nur ungenau einschätzen.

Wie groß ist der Abstand zwischen „selten und manchmal“, „manchmal und oft“, „oft und sehr oft“? Das hängt auch ab von der persönlichen Sichtweise.

Schätzskaalen sollten mindestens zwei oder besser drei Antwortstufen haben (siehe Jürgens/Lissmann 2015, 98). Sie können auch ganz unterschiedlich entworfen sein (siehe z.B. Bühner 2011, 110 ff.).

Für das Rechnen mit Zahlen von Schätzskaalen fehlt rein mathematisch ja meist der Segen. Rein praktisch ist es aber üblich: Zuerst zählt man die angekreuzten Zahlen eines Bereiches zusammen zu einem Summenwert.

Beispiel soziale Entwicklung - angekreuzte Zahlen fett gedruckt:

Das Kind hält sich an die Gruppenregeln.

1 = nie    2 = selten    3 = manchmal    **4** = oft    5 = sehr oft

Es spielt mit anderen Kindern.

1 = nie    2 = selten    3 = manchmal    **4** = oft    5 = sehr oft

Es holt sich Hilfe, wenn es alleine nicht zurechtkommt.

**1** = nie    2 = selten    3 = manchmal    4 = oft    5 = sehr oft

Summenwert:  $4+4+1 = 9$ . Arithmetische Mittel:  $9$  geteilt durch  $3 = 3$ .

Das arithmetische Mittel soll aussagen, wie ein Kind durchschnittlich einzuschätzen ist in diesem Bereich. Für den Vergleich mit altersgleichen Kindern müsste man noch Vergleichsnormen erstellen.

Solch eine quantitative oder zahlenmäßige Auswertung ist sinnvoll trotz des fehlenden mathematischen Segens, wenn man stets daran denkt, dass man hier nur mit Schätzwerten rechnet. Herauskommen können deshalb auch nur ungenaue Orientierungswerte. Mehr erwartet man auch nicht. Und doch: Das klare und vertraute Bild der Zahlen täuscht uns leicht über ihre Ungenauigkeit hinweg.

Man kann auch auswerten, ohne zu rechnen: Die dritte Beobachtungsaufgabe zeigt, dass man das Kind ermutigen sollte, sich Hilfe zu holen, wenn es sie braucht.

## 6. Was ist eine Intervallskala?

Die Schulnoten-Skala wäre nur dann eine Intervallskala, „wenn der Leistungsunterschied zwischen 1 und 2 genauso groß wäre wie der zwischen 3 und 4 oder 5 und 6 usw.“ (Dieterich 1973, 73).

Denn für Intervallskalen gilt: Die Abstände zwischen zwei aufeinander folgenden Zahlen der Skala sind immer gleich groß **und** entsprechen auch gleich großen Merkmalsunterschieden.

Beispiel Celsius-Skala: Der Temperaturunterschied zwischen 10 und 11 Grad Celsius ist genauso groß wie der Unterschied zwischen 75 und 76 Grad Celsius (siehe Freie Universität Berlin, 16, Intervallskala).

Auch viele psychologische Testskalen sind als Intervallskalen entworfen (siehe Clauß/Finze/Partzsch 2011, 11).

Beispiel IQ-Skala: Man nimmt an, dass z.B. der Intelligenzunterschied zwischen IQ 80 und 90 genauso groß ist wie zwischen IQ 110 und 120.

Kann man auch sagen, dass ein IQ von 140 doppelt so hoch ist wie ein IQ von 70? Kann man nicht. Genauso wenig sind 20 Grad Celsius doppelt so warm wie 10 Grad Celsius (siehe Dieterich 1973, 74). Daran ändert auch der Klimawandel nichts.

Das können Sie sich einfach so merken, ohne Erklärung. Denn die ist etwas knifflig.

Für alle, die sie trotzdem oder gerade deshalb wissen möchten:

Intervallskalen haben nur einen willkürlich gesetzten Nullpunkt oder Mittelpunkt, keinen natürlichen (siehe Freie Universität Berlin o. J., 16, Intervallskala).

Beispiel Celsius-Skala: Der willkürliche Nullpunkt ist der Gefrierpunkt des Wassers (siehe Dieterich 1973, 73). Das verhindert Verhältnisvergleiche wie „doppelt so warm“.

Beispiel IQ-Skala: Hier geht man vom Mittelwert aus, also wie weit das Testergebnis einer Person vom Mittelwert der Normstichprobe abweicht - siehe Abschnitt 8.1. Der Mittelwert von 100 ist aber willkürlich gewählt (siehe Freie Universität Berlin o. J., 16, Intervallskala). Das verhindert Verhältnisvergleiche wie „doppelt so intelligent“.

Der grundsätzliche Sachverhalt lässt sich auch mit dem Wettkampf im Weitsprung verdeutlichen: Sprungweiten kann man ja nur dann zueinander ins Verhältnis setzen, wenn man nicht nur die Abstände zwischen den Landepunkten kennt, sondern auch die Abstände der Landepunkte vom Absprungbalken, also vom natürlichen Nullpunkt. Von hier aus gemessen sind 4 Meter wirklich doppelt so weit wie zwei Meter.

Einen natürlichen Nullpunkt kennen Sie z.B. auch von einer Personenwaage. Da sind 140 Kilogramm ja wirklich doppelt so schwer wie 70 Kilogramm, ob man will oder nicht. Skalen

dieser Art sind sogenannte Verhältnisskalen (siehe Freie Universität Berlin o. J., 16, Verhältnisskalen). Das aber nur nebenbei.

Zurück zum IQ: Er sagt nur etwas aus über die quantitativen oder zahlenmäßigen Intelligenzunterschiede zwischen den Menschen (siehe Zimmermann 2011, 4). Doch bestehen zwischen den verschiedenen Stufen der Intelligenz wie unter- oder überdurchschnittlich „auch qualitative Unterschiede, die sich nicht einfach nur durch ‚mehr‘ oder ‚weniger‘ beschreiben lassen. Menschen mit unterschiedlichen Intelligenzquotienten sind in der Regel ‚anders‘ intelligent“ (ebd.). Das gilt besonders für Denkleistungen weit unter oder über dem Durchschnitt (ebd.).

Beispiel Carl Friedrich Gauß, Jahrgang 1777, Mathematiker: Er sagte selbst über sich: „Ich konnte früher rechnen als sprechen“ (Michling 1997, 11, zitiert nach Mittler 2005, 18). Und als eben Dreijähriger soll er einen Fehler entdeckt haben in den Lohnabrechnungen seines Vaters für die Maurergehilfen (siehe Sartorius 1856, 11 f., zitiert nach Mittler 2005, 18). Nicht gerade typisch für Dreijährige!

Hat das getestete Merkmal mindestens Intervallskalenqualität, kann man allerlei berechnen, z.B. das arithmetische Mittel und die sogenannte Standardabweichung (siehe z.B. Wirtz/Nachtigall 2008, 53). Für die Standardabweichung bitte ich aber noch um etwas Geduld. Oder wie es Lehrer „Bömmel“ in der „Feuerzangenbowle“ sagt: „dat krieje mer später“ (Spoerl 2002, 47). Nämlich in Abschnitt 8.1.

## **7. Grundgesamtheit und repräsentative Stichprobe**

Beispiel Intelligenz: Will man die Intelligenz eines Kindes einschätzen, dann benötigt man Vergleichswerte von gleichaltrigen Kindern, die genau diesen Test zuvor gemacht haben. Ein guter Test hat auch gute Vergleichswerte, die schon genannten Vergleichs-Normen (siehe Kreis Borken 2016, 4). Für fehlerfreie Normen müsste man mit dem Test aber vorher alle gleichaltrigen Kinder in Deutschland untersuchen: Die Grundgesamtheit. Das ist jedoch meist zu teuer.

Die Lösung habe ich schon in Abschnitt 2 erwähnt: Man untersucht nur einen Teil der Grundgesamtheit, die Normstichprobe. Aus ihr kann man zwar nur fehlerbehaftete, aber doch geeignete Vergleichs-Normen gewinnen.

Allerdings nur, wenn die Normstichprobe aus einer Teilgruppe gleichaltriger Kinder besteht, die sorgfältig ausgewählt, aktuell und genügend groß ist (ebd.). Die Stichprobe muss also repräsentativ sein für die Grundgesamtheit als ihr verkleinertes, möglichst wahrheitsgetreues Abbild (siehe Clauß/Ebner 1978, 179).

Das ermöglicht eine Zufallsauswahl: Jedes Mitglied der Grundgesamtheit hat die gleiche Chance, Teil der Stichprobe zu werden (siehe Clauß/Ebner 1978, 180). Macht man dabei alles richtig, dann informiert das Ergebnis der Stichprobe nicht nur über die Mitglieder der Stichprobe, sondern erlaubt auch mehr oder minder zuverlässige Aussagen über die Mitglieder der Grundgesamtheit, aus der die Stichprobe stammt (siehe Clauß/Ebner 1978, 18).

### **7.1. Größe der Normstichprobe**

Tests gelten meist für mehrere Altersgruppen, z.B. für Vier-, Fünf- und Sechsjährige. Die Normstichprobe jeder Altersgruppe sollte mindestens 150 Personen umfassen. Bei Altersgruppen, die nach Mädchen und Jungen getrennt sind, gilt die Mindestgröße pro



Geschlecht und Alter (siehe Wyschkon/Esser 2015, 176, Abbildung 1). Das sind dann mindestens 300 Personen pro geschlechtsgetrennter Altersgruppe.

Da die Testentwicklung teuer ist, haben viele Tests eine zu kleine Normstichprobe (siehe Kreis Borken 2016, 4).

## 8. Häufigkeitsverteilung, Rohwerte und Normwerte als standardisierte Werte

Erkennen Sie, was diese Tabelle darstellt?

<b>Noten</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>
Häufigkeiten	4	6	5	3	7	2

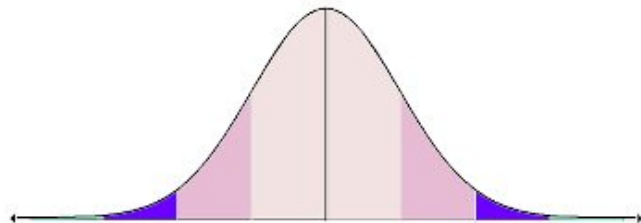
Richtig: Sie ist der Klassen- oder Notenspiegel zu einer Klassenarbeit von 27 Kindern und damit die Häufigkeitsverteilung der Schulnoten innerhalb der Klasse (abgeleitet aus der Rangreihe der Schulnoten Klasse 9a, Eikenbusch/Leuders 2008, 18).

Häufigkeitsverteilungen „geben also an, wie häufig bestimmte Werte vorkommen“ (Eikenbusch/Leuders 2008, 31).

Und wie entstehen Häufigkeitsverteilungen bei Tests? Indem jedes Mitglied der Normstichprobe die Testaufgaben bearbeitet. Man erhält dann sogenannte Rohwerte. Sie ergeben sich unmittelbar aus den Antworten der Testpersonen, z.B. als Anzahl gelöster Aufgaben. Um Rohwerte besser miteinander vergleichen zu können, rechnet man sie häufig um in Standardnormen, z.B. in IQ-Werte (siehe Macha 2012, Rohwert) und/oder in Prozentrangnormen. Diese Normen sind Thema der nächsten Abschnitte.

### 8.1. Normalverteilung und ihre Standardnormen (Normierung)

Der Begriff „Normalverteilung“ ist durchaus treffend, weil diese Häufigkeitsverteilung wirklich schön normal ist. Normal oder normalverteilt bedeutet ungefähr, dass durchschnittliche oder mittlere Messwerte sehr häufig sind und extreme Werte sehr selten (siehe Nguyen-Kim 2021, 202). Das gilt für die meisten Größen in Natur und Alltag, „seien es Körpergrößen, Regenmengen oder Pro-Kopf-Käsekonsum“ (ebd.). Die Normalverteilung hat deshalb eine glockenartige Gestalt:



Beispiel Körpergröße: Es gibt nur wenige sehr kleine und wenige sehr große erwachsene Menschen, die meisten sind mittelgroß. Erfasst man z.B. die Körpergröße aller 30jährigen Frauen des Kreises Borken, so erhält man eine Häufigkeitsverteilung, die sich stark annähert an die Normalverteilung (siehe Kreis Borken 2016, 3). Man spricht auch von der Gauß'schen Glockenkurve und würdigt damit unseren Rechenkünstler Gauß, der die Normalverteilung untersucht hat (siehe Universität Zürich 2023, Abschnitt 5).

Auch bei psychischen Merkmalen erwartet man eine Normalverteilung. So liegt die Intelligenz der meisten Menschen im mittleren Bereich. Nur bei wenigen Menschen liegt sie weit darunter oder darüber (siehe Kreis Borken 2016, 3 f.).

Und was haben die Standardnormen damit zu tun? Viel. Denn man ermittelt sie über die (Standard-) Normalverteilung (siehe Lienert 1961, 320). Man kann sie also nur berechnen, wenn sich die Rohwerte einer Testung normal verteilen (siehe Kleber 1978, 69). Zumindest ungefähr. Denn keine praktisch ermittelte Häufigkeitsverteilung erreicht die ideale Form der Normalverteilung (siehe Kleber 1978, 47).

Weichen die Rohwerte nicht bedeutend ab von der Normalverteilung, rechnet man sie um in Standardnormen (siehe Kipman 2013, 11).

Die gebräuchlichsten Norm- oder Standardskalen wie die IQ-Skala sind alle ähnlich aufgebaut und unterscheiden sich im arithmetischen Mittel und in der Standardabweichung (siehe z.B. Dieterich 1973, 227, Tafel 5).

Die **Standardabweichung** ist ein Streuungsmaß und gibt an, wie die einzelnen Werte einer Häufigkeitsverteilung um den Mittelwert - das arithmetische Mittel - streuen (siehe Macha 2012, Standardabweichung). Sie ist also ein Maß für die Breite einer Häufigkeitsverteilung (siehe Eikenbusch/Leuders 2008, 205).

Verteilen sich die Werte annähernd normal und zieht man die Standardabweichung vom Mittelwert ab und zählt sie zum Mittelwert hinzu, dann umfasst dieser Bereich die mittleren etwa 68 Prozent aller Werte einer Häufigkeitsverteilung.

Beispiel ohne Standardnorm: Die Kinder einer Klassenstufe besuchen pro Woche im Mittel 33.8 Schulstunden. Die Standardabweichung beträgt 2.5 Stunden (siehe Universität Zürich 2023, Abschnitt 3.2.2.). Zieht man nun vom Mittelwert 33.8 die 2.5 Stunden ab und zählt zum Mittelwert die 2.5 Stunden hinzu, dann erhält man den breiten Durchschnittsbereich von 31.3 bis 36.3 Schulstunden. Also besuchen die mittleren etwa 68 Prozent der Kinder 31.3 bis 36.3 Schulstunden pro Woche.

Beispiel mit Standardnorm, IQ-Skala - gilt auch für die EQ-Skala: Sie reicht meist von IQ kleiner als 70 bis größer als 130 (siehe Zimmermann 2011, 3 f.). Und sie hat eine Standardabweichung von 15. Zieht man vom Mittelwert 100 die 15 Punkte ab und zählt zum Mittelwert die 15 Punkte hinzu, dann erhält man den breiten Durchschnittsbereich von IQ 85 bis 115 (siehe Kleber 1978, 50). Hier liegen die mittleren etwa 68 Prozent der Werte der Normstichprobe. Darunter liegen die etwa 16 Prozent der unterdurchschnittlichen und weit unterdurchschnittlichen Werte, darüber die etwa 16 Prozent der überdurchschnittlichen und weit überdurchschnittlichen Werte (siehe Macha 2007a, Grafik).

Beispiel mit Standardnorm, T-Wert-Skala, benannt nach dem Psychologen Terman:

Sie reicht meist von T-Wert kleiner als 30 bis größer als 70, hat einen Mittelwert von 50 und eine Standardabweichung von 10. Das ergibt den breiten Durchschnittsbereich von T-Wert 40 bis 60. Auch hier liegen die mittleren 68 Prozent, darunter die schwächsten 16 Prozent, darüber die besten 16 Prozent (ebd.).

Damit sagen Standardnormen auch aus, wie weit das Ergebnis einer Person vom arithmetischen Mittel der Normstichprobe abweicht, nach unten oder nach oben (siehe Stauche 2008, 2).

Beispiel IQ: Ein IQ von 120 liegt 20 Punkte über dem arithmetischen Mittel von 100 und 5 Punkte über dem breiten Durchschnittsbereich. Er ist also überdurchschnittlich.

## 8.2. Beliebige Verteilung und Prozentrangnormen (Normierung)

Wir wissen nun: Sind die Häufigkeitsverteilungen von Rohwerten so nett, sich der glockenförmigen Normalverteilung anzunähern, dann kann man die Rohwerte in Standardnormen wie den IQ umwandeln. Und wenn sie sich einfach schief verteilen, z.B. bei Schultests? Dann kommt der Prozentrang (PR) ins Spiel:

„Standardnormen können nur aus normalverteilten Rohwerten errechnet werden, Prozentrangnormen hingegen aus jeder beliebigen Verteilung“ (Lienert 1961, 321). Der Prozentrang ist eine reine Häufigkeitsangabe (siehe Fisseni 1997, 115).

Beispiel: Ein Prozentrang von 65 besagt, dass 65 Prozent der Personen der Normstichprobe niedrigere oder gleiche Testleistungen erreichen, 35 Prozent dagegen höhere Leistungen (ebd.).

Es bleibt völlig offen, wie groß die Leistungsunterschiede zwischen Prozenträngen sind.

Beispiel eines sehr schwierigen Intelligenztests: Die Häufigkeitsverteilung der Rohwerte oder Punkte ist schief. Möglich sind 50 Punkte. Die Ergebnisse von vier Personen:

**Anna** erreicht 10 Punkte und Prozentrang 10, **Benn** 15 Punkte und Prozentrang 15. Der Leistungs-Unterschied von 5 Punkten entspricht einem Unterschied von 5 Prozenträngen.

**Clara** erreicht 40 Punkte und Prozentrang 80, **Doris** 45 Punkte und Prozentrang 95. Der Leistungs-Unterschied beträgt wieder 5 Punkte, das ist aber nun ein Unterschied von 15 Prozenträngen (siehe Bühner 2011, 264 f.).

Das heißt: Prozentrang-Unterschiede kann man nicht direkt in Leistungs-Unterschiede übertragen (siehe Bühner 2011, 265). Mit einer Prozentrangskala kann man nur sagen, „ob jemand besser war als eine andere Person, aber nicht, um wie viel“ (ebd.). Deshalb darf man auch mit Prozenträngen nicht rechnen und kein arithmetisches Mittel berechnen aus zwei oder mehr Prozenträngen (siehe Kleber 1978, 70).

Der Mittelwert ist hier der **Median (Zentralwert)**. Nehmen wir wieder unser Beispiel mit dem Notenspiegel:

<b>Noten</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>
Häufigkeiten	4	6	5	3	7	2

Man bestimmt den Median dieser Häufigkeitsverteilung, indem man alle Noten nach ihrem Rang ordnet und den Wert in der Mitte der Rangreihe auswählt (siehe Eikenbusch/Leuders 2008, 18):

1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 4, 4, 4, 5, 5, 5, 5, 5, 5, 5, 5, 6, 6

Denn für den Median gilt, dass genauso viele Werte unter wie über ihm liegen (siehe Universität Zürich 2023, Abschnitt 3.1.2.). Also ist der Median als Mittelwert hier die unterstrichene Note 3 (siehe Eikenbusch/Leuders 2008, 18).

Auf einer Skala ist der Median der Leistungswert mit dem Prozentrang 50 (siehe Kleber 1978, 256). Ein Prozentrang von 50 besagt, dass 50 Prozent der Personen der Normstichprobe niedrigere oder gleiche Testleistungen erreichen. Auf der IQ-Skala ist das der IQ 100, auf der T-Wert-Skala der T-Wert 50.

Die Prozentrangskala hat einen schmalen Durchschnittsbereich: Die Prozentränge 25 bis 75 (siehe Kleber 1978, 70). Hier liegen die mittleren 50 Prozent der Werte (siehe Kleber 1978, 256 f.; siehe Clauß/Ebner 1978, 94 f.).

Prozentränge kleiner als 25 sind unterdurchschnittlich, Prozentränge größer als 75 sind überdurchschnittlich. Die Prozentränge 25 bis 75 entsprechen IQ 90 bis 110, die Prozentränge 16 bis 84 entsprechen IQ 85 bis 115.

### **8.3. Normtabellen**

Die Rohwerte eines Tests werden mit den zugehörigen Standardwerten und/oder Prozenträngen dargestellt in einer Normtabelle. So kann jede Fachkraft zu einem erreichten Rohwert schnell den zugehörigen Standardwert und/oder Prozentrang ablesen. Eine beispielhafte Normtabelle mit Rohwerten, T-Werten und Prozenträngen finden Sie im Internet bei Trolldenier (o. J., Abschnitt 2.1.5., Tabelle 10.3).

### **8.4. Alter der Normen**

Altersbezogene Angaben für Kinder im Kindergarten sollten höchstens 10 bis 15 Jahre alt sein. Warum? Weil heute die Lebensumstände von Kindern im Vergleich zu vergangenen Jahrzehnten z.B. geprägt sind durch ein verändertes Spiel- und Freizeitverhalten, eine andere Erziehung und bessere frühkindliche Bildung (siehe Macha/Petermann 2017, 143). Deshalb empfiehlt sich schon ab dem Vorschulalter eine Neunormierung von Entwicklungstests in Abständen von zehn Jahren (siehe Macha/Petermann 2013, 188).

Auch sonst sollten die Normen von Leistungstests nach 10 Jahren überprüft werden (siehe z.B. Perleth o. J., 5). Oder spätestens nach 15 Jahren, wie bei Intelligenztests (siehe Kanaya et al. 2005, zitiert nach Petermann/Daseking 2015, 107).

Es gilt immer das Jahr, in dem der Test normiert wurde, nicht das Jahr seiner Veröffentlichung. Wegen der schon erwähnten teuren Testentwicklung haben viele Verfahren auch veraltete Normen (siehe Kreis Borken 2016, 4).

## **9. Messfehler oder Ungenauigkeit unserer Test- und Beobachtungsverfahren**

„Alle Messinstrumente sind mit einem Messfehler behaftet ...“ (Zimmermann 2011, 15). Das ist unvermeidlich und gilt natürlich auch für psychologische Tests.

Frage: Wodurch entstehen diese Messfehler eigentlich? Antwort: Durch Fehler bei der Entwicklung, Durchführung und Auswertung der Tests (siehe Schmidt-Atzert/Amelang 2012, 43). Mit der Testentwicklung hatte ich nie zu tun, deshalb lasse ich sie aus.

Zur Durchführung: Es fängt schon damit an, dass die Testperson einen guten Tag erwischt haben kann oder einen schlechten. Besonders bei Kindern in den ersten drei Lebensjahren hängt die Leistung stark ab von Tagesform und gegenwärtigem Befinden (siehe Wiefel et al. 2007, zitiert nach Koglin/Hallmann/Petermann 2011, 5).

„Auch emotionale Probleme können eine Rolle spielen, etwa der Streit der Eltern am Morgen, Prüfungsangst, Blockaden unter Zeitdruck oder fehlende Motivation“ (Eikenbusch/Leuders 2008, 46).

Man kann auch nie voraussehen, ob und wie sich ein Kind beeinflussen lässt durch Merkmale der Testsituation. Das können alle möglichen Merkmale der Fachkraft sein wie Geschlecht, Alter, Stimme (siehe Schmidt-Atzert/Amelang 2012, 44), eine modische Brille oder ein Bart.

Es können auch die Lichtverhältnisse sein, die Raumtemperatur, Geräusche von außen (ebd.), ein Bild und viele andere Kleinigkeiten. Erfahrungsgemäß machen die meisten Kinder aber gut mit.

Zur Testauswertung: Da unterlaufen sogar Fachleuten bisweilen Fehler. Damit Sie nicht denken, dass ich hier nur aus eigener Erfahrung spreche, bemühe ich auch hier die Literatur: So können Fachleute z.B. Punkte falsch auszählen oder in einer Normtabelle zu einem Rohwert den falschen Standardwert ablesen (ebd.).

Die gute Nachricht: Man kann den durchschnittlichen Messfehler eines Tests ermitteln; er ist für die Interpretation oder Auslegung unerlässlich und steht daher immer im Manual oder Handbuch des Tests (siehe Zimmermann 2011, 15).

Mit seiner Hilfe kann man einen Wertebereich um den gemessenen Wert eingrenzen, innerhalb dessen der wahre oder messfehlerfreie Wert liegt, z.B. mit einer Sicherheit von 68, 80 oder 95 Prozent (siehe Macha 2012, Konfidenzintervall).

Es ist eine „Unsicherheits-Region um den gemessenen Wert“ (Schilling 2012, 185). Allgemein spricht man von Vertrauensbereich oder Konfidenzintervall.

**Beispiel Intelligenztest:** Festgestellt wurde ein IQ von 90. Man wählt eine Sicherheit von 95 Prozent, das entspricht 5 Irrtümern in 100 Fällen. Der durchschnittliche Messfehler beträgt dann bei diesem Test plus/minus 9 IQ-Punkte. Damit liegt der wahre IQ im Vertrauensbereich zwischen IQ 81 und 99 (siehe Zimmermann 2011, 15). Das ist eine unterdurchschnittliche bis durchschnittliche Leistung.

Die Berücksichtigung des Messfehlers ist extrem wichtig, wenn vom Testergebnis Entscheidungen zur Schullaufbahn abhängen (ebd.): Erweist sich z.B. der Wechsel in eine andere Schulform als Fehler, dann kann man das nur schwer berichtigen.

**Schultests:** Sie haben oft schiefe Rohwertverteilungen mit Prozentrangnormen. Nachteil: Bei der Prozentrangskala kann man die Ungenauigkeit eines Tests nur bis zu einem gewissen Grade berücksichtigen, und zwar mit einem Prozentrangband (siehe Kleber 1978, 70 f.). In diesem Band liegt der wahre Leistungswert einer Person mit einer Sicherheit von 68 Prozent (siehe Kleber 1978, 85 f.).

Beispiel: Bei einem Rechentest liegt die wahre Leistung von Felix mit 68 Prozent Sicherheit zwischen Prozentrang 20 und 36.

Wie das Leben so spielt, gibt es neben den schon genannten zufälligen oder unsystematischen Fehlern noch solche, die nicht zufällig sind, sondern systematisch. Bei ihnen ist selbst der schönste Vertrauensbereich für die Katz (denn die Klassische Testtheorie, die ich hier anspreche, berücksichtigt nur unsystematische Messfehler - siehe Bühner 2011, 42).

Beispiele für systematische Fehler: Sportlerinnen und Sportler können einen Teil ihrer Leistungen unter Dopingeinfluss vollbracht haben. Und eine Testperson kann vorab schon einen Teil der Aufgaben und Lösungen kennen (siehe Krumm/Schmidt-Atzert/Amelang 2021, 50).

**Beobachtungsbögen und Schätzskalen:** Bei allen Verhaltensbeobachtungen sind Fehler oder Verzerrungstendenzen möglich (siehe Fisseni 1997, 199 ff.). Ihnen ist gemeinsam, dass man Sachverhalte nicht feststellt, sondern nur annimmt aufgrund unbegründeter Vor-Urteile (siehe Fisseni 1997, 209). Denn unsere Wahrnehmung ist beileibe nicht so unbestechlich wie eine

Kamera oder ein Tonband. Nicht umsonst gibt es im Profisport den Videobeweis. Und nicht umsonst heißt es, dass Liebe blind macht.

Fehlen klare Richtlinien, wie ein Verhalten zu beurteilen ist, dann muss man davon ausgehen, dass die Beurteilung stark abhängt von der beurteilenden Person. Schon Sympathie oder Antipathie können folgeschwer sein. Finden wir jemanden sympathisch, dann sind wir eher geneigt, Positives bei ihm wahrzunehmen; können wir jemanden nicht leiden, sehen wir ihn eher negativ (siehe Flender/Tröster 2005, 7).

Weitere typische Fehler finden Sie im Internet z.B. bei Macha (2012, Beurteilungsfehler).

## 10. Qualitätsmerkmale von Tests: Hauptgütekriterien

Damit Messfehler möglichst klein bleiben, muss ein Test drei Hauptgütekriterien erfüllen: Er soll objektiv, zuverlässig und gültig sein (siehe Lienert 1961, 12). Das gilt auch für unsere Beobachtungsbögen.

### 10.1. Objektivität = Unabhängigkeit der Messung

Die Testergebnisse sollten möglichst unabhängig sein von den durchführenden und auswertenden Personen. Deshalb sollten verschiedene Personen zu weitgehend übereinstimmenden Ergebnissen kommen (siehe Zimmermann 2011, 15).

Beispiel Beobachtungsbogen: Er hat eine hohe Objektivität, wenn zwei Fachkräfte mit einem Bogen ein bestimmtes Kind beobachten und unabhängig voneinander zum gleichen Ergebnis kommen (siehe Flender/Tröster 2005, 10).

Frage: Womit kann man die Objektivität eines Verfahrens herstellen? Antwort: Mit seiner Standardisierung (siehe Macha 2012, Objektivität, Standardisierung).

Standardisierungen oder Vereinheitlichungen sind alltäglich: So ist ein DIN A 4 Blatt immer gleich groß und ein Kochrezept beschreibt ein einheitliches Vorgehen.

Voll standardisiert sind **Entwicklungs-, Intelligenz- und Schultests**. Schon das Testmaterial ist einheitlich. Das sind z.B. Puzzles, Bilder zum Legen von Geschichten und Hefte mit Rechtschreib- oder Rechenaufgaben. Nur dieses Material darf verwendet werden. Warum? Weil schon eine kleine Veränderung des Materials eine Aufgabe schwerer oder leichter machen kann (siehe Macha/Petermann 2013, 188).

Auch das Verhalten der Fachkraft und die Testsituation sind festgelegt. Jedes Kind, das einen bestimmten Test macht, „bekommt die gleichen Aufgaben in der gleichen Zeit zu gleichen Bedingungen vorgelegt. Die Antworten werden für jedes Kind nach den gleichen Regeln ausgewertet“ (Kreis Borken 2016, 4).

**Beobachtungsbögen:** Um Verzerrungen bei der Beobachtung möglichst klein zu halten, sind viele Bögen immerhin teil- oder halbstandardisiert.

Beispiel körpermotorische Entwicklung:

Das Kind steigt eine Treppe hinauf, selbstständig und im Wechselschritt.

1 = noch nicht möglich 2 = eher nicht möglich 3 = eher möglich 4 = möglich

Dies ist eine teilstandardisierte Beobachtungsaufgabe mit vorgegebenen Antwortstufen. Für eine volle Standardisierung fehlen zumindest noch Angaben zur Anzahl und Höhe der Stufen und ob sich das Kind festhalten darf am Geländer.

Bei anderen Aufgaben geht es um einen breiten Verhaltensbereich mit verschiedenen Situationen. Sie sind deshalb nur allgemein gefasst und wenig standardisiert.

Beispiel emotionale Entwicklung: Kann angemessen mit Enttäuschungen umgehen.

1 = trifft gar nicht zu 2 = trifft wenig zu 3 = trifft überwiegend zu 4 = trifft zu

Hier bleibt offen, welche Enttäuschungen gemeint sind und wann das Verhalten des Kindes als angemessen zu bewerten ist.

„Gerade für die Durchführung einer Beobachtung im Kindergartenalltag sind aber Einschränkungen der Objektivität zu erwarten. So ist eine Standardisierung nur schwer möglich, wenn die Aufgabe eine längerfristige Beobachtung des Kindes in verschiedenen Alltagssituationen erfordert“ (Flender 2005, 38). Es ist dann nicht immer möglich, genaue Kriterien oder Richtlinien zu Durchführung und Auswertung der Aufgaben zu benennen (siehe Flender 2005, 120). Das begünstigt Fehler.

Dafür beschert uns die längerfristige Beobachtung eines Kindes im Alltag viele lebensechte Informationen, die man mit Tests nicht erfassen kann.

### **10.2. Reliabilität = Zuverlässigkeit, Genauigkeit**

Ein Test ist zuverlässig, wenn er das Merkmal, das er misst, genau misst (siehe Lienert 1961, 13). Häufig bestimmt man die Wiederholungs-Zuverlässigkeit (Re-Test-Reliabilität): Sie gibt an, inwieweit man bei einer Test-Wiederholung zu ähnlichen Ergebnissen kommt (siehe Kreis Borken 2016, 4). Oder bei wiederholter Beobachtung.

Beispiel Beobachtungsbogen: Jeweils ein und dieselbe Fachkraft beobachtet ein Kind im Abstand von wenigen Wochen zweimal mit demselben Bogen. Sind beide Ergebnisse gleich oder ähnlich, ist der Bogen zuverlässig.

### **10.3. Validität = Gültigkeit, Richtigkeit**

Die Validität „zeigt, wie gut ein Test das erfasst, was er messen soll“ (Kipman 2013, 13). Denn ein Test kann zwar genau messen, aber doch nicht das, was er messen soll.

Es gibt mehrere Arten von Validität (siehe z.B. Kipmann 2013, 13 f.). Eine davon ist die kriterienbezogene Validität: Hier geht es darum, wie stark ein Test übereinstimmt mit einem Außenkriterium oder Außenmaß, z.B. mit einem anderen Test oder einer Fremdeinschätzung (siehe Macha 2012, Validität, kriterienbezogene). Das gilt ebenso für Beobachtungsbögen.

Beispiel Beobachtungsbogen: Fachleute schätzen die motorische Entwicklung von Kindern mit einem Bogen ein. Bei einzelnen Kindern zeigen sich auffällige Ergebnisse. Der Bogen ist gültig, wenn die Ergebnisse durch andere Fachleute bestätigt werden können; das geht mit motorischen Tests oder kinderärztlichen Beobachtungen (siehe Flender/Tröster 2005, 11).

**Screenings** sind wegen ihrer Kürze meist weniger zuverlässig und gültig als Tests. Weitere Hinweise finden Sie z.B. bei Macha/Proske/Petermann (2005, 152).

## **11. Einige Möglichkeiten, sich über Tests zu informieren**

Die Bedeutung eines Tests ist nicht leicht zu beurteilen. Denn es kommt auch darauf an, was der Test in welcher Situation zur Lösung einer Fragestellung beitragen kann (siehe Bühner 2011, 81). So kann es mitunter sinnvoll sein, einem „schwachen“ Test Aufgaben oder Aufgabenreihen zu entnehmen für bestimmte diagnostische Zwecke (siehe Perleth o. J., 5).

Informationen zu einem Test stehen im jeweiligen Manual oder Handbuch.

Es gibt auch ein Verzeichnis von Fachzeitschriften mit Testbesprechungen. Sie finden es unter: [https://www.psyndex.de/pub/tests/verz\\_teil5.pdf](https://www.psyndex.de/pub/tests/verz_teil5.pdf)

Macha (2012) erklärt sehr viele Fachbegriffe kurz und gut im Internet.

## **12. Ein Kind wird oder wurde getestet – einige Informationen und Hinweise**

Zuallererst muss ärztlich untersucht sein, ob das Kind richtig hört und sieht. Brille oder Hörhilfe muss es beim Test tragen.

Es sollte ausgeschlafen und gesund sein und weder hungrig noch durstig. Es sollte auch aus keiner anstrengenden oder belastenden Situation kommen. Klein- und Vorschulkinder testet man natürlich nur vormittags. Das gilt normalerweise auch für Grundschulkindern, die dann schulfrei haben sollten. Ältere Schulkinder kann man auch am frühen Nachmittag testen, nachdem sie sich von der Schule erholt haben. Intelligenztests habe ich aber auch bei ihnen nur vormittags durchgeführt.

**Wie können Eltern ihr Kind auf einen Test vorbereiten?** Das hängt davon ab, wie alt das Kind ist und was es verstehen kann (siehe Kreis Borken 2016, 6). Bei einem kleinen Kind genügt es, wenn die Eltern ihm z.B. sagen: „Wir möchten gerne wissen, was du schon alles weißt und kannst; wir kennen auch jemanden, der das herausfinden kann; und es gibt dort Aufgaben und Rätsel, die extra für Kinder sind und den meisten Kindern sehr viel Spaß machen“ (siehe Schuhmacher o. J.).

Bei Schulkindern ist die Sache klarer. Denn die meisten Kinder kennen ihre Lernprobleme. So wird kein Schulkind verwundert sein über einen Rechtschreibtest zur Abklärung seiner Rechtschreibprobleme (siehe Kreis Borken 2016, 6).

Dennoch kann ein passender Vergleich hilfreich sein, z.B.: „Es ist wie beim sportlichen Training. Es wird geschaut, was du schon kannst und was du noch trainieren solltest.“

Auch Fachleute drücken sich immer kindgemäß aus. Beispielsweise merken selbst jüngere Kinder bei Intelligenztests schnell, dass es um ihre Leistungsfähigkeit geht. Die Fachkraft betont dann aber nicht die Begabungsfrage (ebd.). Stattdessen wird sie z.B. sagen: „Ich möchte schauen, was du gut und weniger gut kannst“ (ebd.).

**Sollte ein Elternteil bei der Testung dabei sein?** Ist das Kind schon vier Jahre alt, dann in der Regel nicht mehr (siehe z.B. Esser/Petermann 2010, zitiert nach Wyschkon/Esser 2015, 166). Es könnte sonst abgelenkt sein und elterlichen Druck verspüren (siehe Wyschkon/Esser 2015, 166). Eltern sind auch oft nicht passiv genug und sagen Sätze wie „Zu Hause kann er das aber schon!“ (Macha 2007b).

Ist es nötig, dass ein Elternteil bei einem älteren Kind dabei bleibt, sollte er abseits sitzen ohne direkten Blickkontakt zum Kind.

**Nach der Testung:** Die Fachkraft hat das Kind schon während der Diagnostik bestärkt. Auch die Eltern sollten ihrem Kind sagen, dass es seine Arbeit gut gemacht hat, unabhängig vom zu erwartenden Ergebnis (siehe Andrich o. J.). Kinder sprechen meist sehr gerne über die Dinge, die sie während der Testung lösen konnten, und sind sehr stolz darauf. Die Eltern sollten ihrem Kind dann einfach nur gut zuhören (ebd.). Ungünstig sind wertende Aussagen wie „na das kannst du aber eigentlich besser“ (ebd.). Sie können dem Kind weitere Termine



erschweren. Manchmal brauchen Kinder noch etwas Zeit, bevor sie von der Diagnostik erzählen. Die sollten sie haben (ebd.).

**Besprechung der Testergebnisse:** Fachleute besprechen ihre Ergebnisse bei Kleinkindern, Kindergarten- und Grundschulkindern in der Regel nur mit den Eltern. Ältere Schulkinder können an der Besprechung teilnehmen. Dabei wählen Fachleute ihre Worte gut. Rückmeldungen an die Kinder durch Fachleute oder Eltern sollten immer die persönlichen Stärken der Kinder betonen.

**Allgemeine Entwicklungstests:** Der Gesamt-EQ gilt heute als zu allgemein (siehe Macha 2007c, Abschnitt 2.4.). Zudem hat er ja auch einen Messfehler. Deshalb ist es ratsam, den Eltern allenfalls den Vertrauensbereich zu nennen, in dem der wahre Gesamt-EQ liegt mit einer gewählten Sicherheit. Das geht auch ohne Zahlen.

Beispiel: „Nach den Aufgaben dieses Entwicklungstests ist Ihr Kind derzeit insgesamt durchschnittlich entwickelt, und zwar mit großer Sicherheit.“

Viel wichtiger ist das Entwicklungsprofil, das ja Stärken und Schwächen aufzeigen kann. Für eine Förderempfehlung muss man aber meist auch noch Entwicklungsrisiken und Schutzfaktoren berücksichtigen (siehe Macha/Petermann 2017, 149 f.). Entwicklungsrisiken sind z.B. Erkrankungen des Kindes oder der Eltern und ein ungünstiges Erziehungsverhalten (siehe Macha/Petermann 2016, 101). Schutzfaktoren sind z.B. ein günstiges Erziehungsverhalten und Hobbys wie Sport oder Musik. Sind die Gesamtbedingungen gut, dann kann man z.B. bei einer leichten Entwicklungsverzögerung gegebenenfalls noch abwarten, bevor man weitere Maßnahmen erwägt (siehe Macha/Petermann 2017, 137 f.).

Ganz wichtig ist also immer, was die Eltern aus Vorgeschichte und Alltag berichten. Hinzu kommen die Einschätzungen des Kindergartens und ärztliche Befunde, wenn die Eltern dies möchten.

Bei Macha/Petermann (2017, 150 f.) finden Sie grundsätzliche Entscheidungsregeln zu der Frage, was Testergebnisse über einen Förder- oder Therapiebedarf aussagen.

**Allgemeine Intelligenztests:** „Die Intelligenz eines Kindes mit einer abschließenden Zahl (z.B. Gesamt-IQ) darzustellen, ist als Gedanke verwegen, kommt aber in der Praxis dem Wunsch nach Vereinfachung und Orientierung entgegen“ (Joel 2018, 198). Dieser Wunsch ist natürlich verständlich.

Dennoch ist es geboten, mit den Ergebnissen von Intelligenztests besonders vorsichtig umzugehen. Zumal die Öffentlichkeit der Intelligenz zu viel Gewicht beimisst (siehe Zimmermann 2011, 19). So ist das unbedachte Nennen von IQ-Werten ein Kunstfehler, der weitreichende Folgen für die Betroffenen haben kann. Empfehlenswert sind auch hier altersangemessene Rückmeldungen an das Kind, die dazu beitragen, gewünschte Änderungen vorzubereiten (siehe Kreis Borken 2016, 6): „Das Ergebnis deines Wissenstests war sehr gut - der Rechtschreibtest war dagegen ausreichend“ (ebd.).

Viel wichtiger als der Gesamt-IQ ist wiederum das Intelligenzprofil. Auch hier sollten nur die Vertrauensbereiche genannt werden.

Zudem gilt für alle Leistungstests der Grundsatz: Ein positives Testergebnis ist eindeutig. Beispielsweise hat die Person bei einem Intelligenztest durch ihre Testleistung gezeigt, dass sie mindestens über so viel Intelligenz verfügt, wie sie im Test bewiesen hat. Dagegen kann ein negatives Testergebnis viele Gründe gehabt haben, so dass man daraus nicht sicher auf

mangelnde Intelligenz schließen kann. Die Person sollte dann noch mindestens einen weiteren Intelligenztest machen, möglichst an einem anderen Tag und mit einer anderen Fachkraft (siehe Zimmermann 2011, 14 f.). Es darf aber nicht darum gehen, so lange zu testen, bis man ein gewünschtes Ergebnis hat.

Natürlich dürfen die Eltern die Test-Unterlagen einsehen. Sie sollten sich dann alle Ergebnisse fachlich genau erklären lassen.

**Schultests:** Ihre Ergebnisse ergänzt man immer durch die Einschätzung der Lehrkräfte, wenn die Eltern dies wünschen.

### 13. Schlussbemerkungen

„Alle, auch die am besten testpsychologisch abgesicherten Ergebnisse der Diagnostik stellen nur Schätz- oder Näherungswerte dar“ (Kleber 1978, 19 f.).

Unpassend ist deshalb eine übertriebene Testgläubigkeit (siehe Schilling 2012, 185). Eine übertriebene Testkritik aber auch. Warum? Weil es häufig nichts Besseres gibt. Ein Test erlaubt oft einen raschen und umfassenden Einblick in die Ausprägung von Merkmalen, und das vergleichsweise objektiv. Nimmt man nur Beobachtungen, kann man sich unnötige Verzerrungen einhandeln. Zudem berücksichtigen gewissenhafte Fachleute die Grenzen ihrer Verfahren (siehe Kreis Borken 2016, 5).

Kurzum: Tests und Beobachtungsbögen sind wichtige Bausteine bei der Einschätzung von Entwicklung, Begabung und Schulleistung, wenn man ihre Möglichkeiten richtig einschätzt und ihre Ergebnisse ergänzt durch weitere Informationen.

#### Hinweise zum Text:

Dieser Text ist eine stark überarbeitete, erweiterte und aufgelockerte Fassung von: Herold, F. (2020): Was Sie schon immer über die statistischen Hintergründe von Verfahren zur Entwicklungseinschätzung wissen wollten. In: heilpädagogik.de – Fachzeitschrift des Berufs- und Fachverbandes Heilpädagogik E.V., 35. Jg., Heft 4, Online-Teil.

[https://bhponline.de/download/Zeitung/online-inhalte/2020-04/01\\_6\\_Fachartikel-Statistische-Hintergruende.pdf](https://bhponline.de/download/Zeitung/online-inhalte/2020-04/01_6_Fachartikel-Statistische-Hintergruende.pdf)

### Quellenverzeichnis

Andrich, J. (o. J.): Psychologische Diagnostik, Kinder/Jugendliche: Vorbereitung auf die Diagnostik - Nach der Diagnostik.

<https://www.psychologisediagnostik.com/kinder-jugendliche/vorbereitung/>  
(Zugriff am 26.02.2023)

Bartel, H. (1971): Statistik I für Psychologen, Pädagogen und Sozialwissenschaftler. Als Studienbegleiter, zum Selbststudium und als Orientierungshilfe in der empirischen Forschung. Stuttgart: Fischer.

Bühner, M. (2011): Einführung in die Test- und Fragebogenkonstruktion. München, Harlow, Amsterdam, Madrid, Boston, San Francisco, Don Mills, Mexico City, Sydney: Pearson, 3., aktualisierte und erweiterte Auflage.

Clauß, G., Ebner, H. (1978): Grundlagen der Statistik. Für Psychologen, Pädagogen und Soziologen. Berlin: Volk und Wissen, 6. Auflage.

Clauß, G., Finze, F.-R., Partzsch, L. (2011): Grundlagen der Statistik. Für Soziologen, Pädagogen, Psychologen und Mediziner. Frankfurt am Main: Deutscher, 6., korrigierte Auflage.

Dieterich, R. (1973): Psychodiagnostik. Grundlagen und Probleme. München, Basel: Reinhardt.

Dinter, R. (2011): Fehlerrechnung für Einsteiger. Eine beispielorientierte Einführung für Studierende der TUHH. Institut für Angewandte Physik, Universität Hamburg.

<https://www2.physnet.uni-hamburg.de/TUHH/Versuchsanleitung/Fehlerrechnung.pdf>  
(Zugriff am 30.09.2023)

Eikenbusch, G., Leuders, T. (Hrsg.) (2008): Lehrer-Kursbuch Statistik. Alles über Daten und Zahlen im Schulalltag. Berlin: Cornelsen, 5. Auflage.

Fisseni, H.-J. (1997): Lehrbuch der psychologischen Diagnostik. Mit Hinweisen zur Intervention. Göttingen, Bern, Toronto, Seattle: Hogrefe, 2., überarbeitete und erweiterte Auflage.

Flender, J. (2005): Früherkennung von Entwicklungsstörungen durch Erzieherinnen: Überprüfung der Gütekriterien des Dortmunder Entwicklungsscreening für den Kindergarten (DESK 3-6). Dortmund: Dissertation.

Flender, J., Tröster, H. (2005): Beobachtungsbögen (1). Wie finden Sie den richtigen für sich? In: kindergarten heute, 35. Jg., Nr. 9, 6-12.

Freie Universität Berlin (o. J.): Themenblock 2: Lernerfolgskontrollen – C) Statistische Auswertung.

[https://www.ewi-psy.fu-berlin.de/erziehungswissenschaft/arbeitsbereiche/lernpsych/Paedagogische\\_Diagnostik/BL\\_Statistische\\_Auswertung/000/index.html](https://www.ewi-psy.fu-berlin.de/erziehungswissenschaft/arbeitsbereiche/lernpsych/Paedagogische_Diagnostik/BL_Statistische_Auswertung/000/index.html)  
(Zugriff am 30.09.2023)

Joel, T. (2018): Intelligenzdiagnostik mit geflüchteten Kindern und Jugendlichen. In: Zeitschrift für Heilpädagogik, 69. Jg., Nr. 5, 196-206.

Jürgens, E., Lissmann, U. (2015): Pädagogische Diagnostik. Grundlagen und Methoden der Leistungsbeurteilung in der Schule. Weinheim, Basel: Beltz.

Kipman, U. (2013): Psychologische Diagnostik moderierender Persönlichkeitsmerkmale bei Kindern und Jugendlichen. Broschüre. Salzburg: Eigenverlag, Österreichisches Zentrum für Begabtenförderung und Begabungsforschung (ÖZBF).

Kleber, Ed. W. (1978): Lehrbuch der sonderpädagogischen Diagnostik. Eine Einführung in die Grundlagen diagnostischer Informationserhebung zum Zwecke pädagogischer Beratung für Sonderpädagogen, Lehrer und Berater im schulischen Bereich. Berlin: Marhold, 3., völlig neubearbeitete und erweiterte Auflage von „Grundlagen sonderpädagogischer Diagnostik“.

Koglin, U., Hallmann, A., Petermann, F. (2011): Entwicklungsdiagnostische Verfahren für die Arbeit mit Kindern in den ersten drei Lebensjahren. KiTa Fachtexte.

[https://www.kitafachtexte.de/fileadmin/Redaktion/Publikationen/KiTaFT\\_KoglHallPeterm\\_2011.pdf](https://www.kitafachtexte.de/fileadmin/Redaktion/Publikationen/KiTaFT_KoglHallPeterm_2011.pdf)  
(Zugriff am 26.02.2022).

Kreis Borken, Regionale Schulberatungsstelle (Hrsg.) (2016): Psychologisches Testen. Informationen für Eltern und Lehrkräfte. Broschüre. Borken.

[https://rsb-borken.de/fileadmin/Ressourcen/schulpsychologie/2016-Broschuere\\_-\\_Psychologisches\\_Testen.pdf](https://rsb-borken.de/fileadmin/Ressourcen/schulpsychologie/2016-Broschuere_-_Psychologisches_Testen.pdf)  
(Zugriff am 30.09.2023)

Krumm, S., Schmidt-Atzert, L., Amelang, M. (2021): Grundlagen diagnostischer Verfahren. In:

Schmidt-Atzert, L., Krumm, S., Amelang, M. (Hrsg.) (2021): Psychologische Diagnostik. Berlin, Heidelberg: Springer, 6., vollständig überarbeitete Auflage, 39 - 207.

- Lienert, G. A. (1961): Testaufbau und Testanalyse. Weinheim: Beltz.
- Macha, T. (2007a): Entwicklungsdiagnostik - Standardwerte.  
<http://entwicklungsdiagnostik.de/standardwerte.html>  
(Zugriff am 27.02.2020)
- Macha, T. (2007b): Entwicklungsdiagnostik - Empfehlung.  
<http://entwicklungsdiagnostik.de/empfehlung.html>  
(Zugriff am 27.01.2023)
- Macha, T. (2007c): Entwicklungsdiagnostik - Grundlagen.  
<http://entwicklungsdiagnostik.de/grundlagen.html>  
(Zugriff am 04.03.2022)
- Macha, T. (2009): Grundintelligenztest Skala 1  
[http://entwicklungsdiagnostik.de/cft\\_1.html](http://entwicklungsdiagnostik.de/cft_1.html)  
(Zugriff am 14.10.2023)
- Macha, T. (2012): Entwicklungsdiagnostik - Glossar.  
<http://entwicklungsdiagnostik.de/glossar.html>  
(Zugriff am 24.10.2019).
- Macha, T., Petermann, F. (2013): Objektivität von Entwicklungstests. Zur Standardisierung der entwicklungsdiagnostischen Befunderhebung. In: Diagnostica, 59. Jg., Nr. 4, 183-191.
- Macha, T., Petermann, F. (2016): Mein Kind und ich. Die ersten Lebensjahre. Bern: Hogrefe.
- Macha, T., Petermann, F. (2017): Entwicklungsdiagnostische Verfahren: Ressourcen- und Risikoerkennung. In: Petermann, F., Wiedebusch, S. (Hrsg.) (2017): Praxishandbuch Kindergarten. Entwicklung von Kindern verstehen und fördern. Göttingen: Hogrefe, 133-152.
- Macha, T., Proske, A., Petermann, F. (2005): Validität von Entwicklungstests. In: Kindheit und Entwicklung, 14. Jg., Nr. 3, 150-162.
- Mittler, E. (Hrsg.) (2005): „Wie der Blitz einschlägt, hat sich das Räthsel gelöst“. Carl Friedrich Gauß in Göttingen. Katalog. Göttingen: Niedersächsische Staats- und Universitätsbibliothek Göttingen.
- Nguyen-Kim, M. T. (2021): Die kleinste gemeinsame Wirklichkeit. Wahr, falsch, plausibel? Die größten Streitfragen wissenschaftlich geprüft. München: Droemer.
- Pauen, S. (2017): Milestones of Normal Development in Early Years (MONDEY). In: Petermann, F., Wiedebusch, S. (Hrsg.) (2017): Praxishandbuch Kindergarten. Entwicklung von Kindern verstehen und fördern. Göttingen: Hogrefe, 172 - 193.
- Perleth, C. (o. J.): Grundlagenwissen der Diagnostik für Nicht-Psychologen. Universität Rostock.  
<https://docplayer.org/12513545-Grundlagenwissen-der-diagnostik-fuer-nichtpsychologen.html>  
(Zugriff am 08.02.2022).
- Petermann, F., Daseking, M. (2015): Diagnostische Erhebungsverfahren. Göttingen, Bern, Wien, Paris, Oxford, Prag, Toronto, Boston, Amsterdam, Kopenhagen, Stockholm, Florenz, Helsinki, Sao Paulo: Hogrefe.
- Schilling, F. (2012): Diagnose und Befund bei Kindern mit Entwicklungsstörungen der motorischen Funktionen. In: praxis ergotherapie, 25. Jg., Nr. 6, 184-189.
- Schmidt-Atzert, L., Amelang, M. (2012): Psychologische Diagnostik. Berlin, Heidelberg: Springer, 5., vollständig überarbeitete und erweiterte Auflage.

Schmidt-Atzert, L., Krumm, S., Amelang, M. (Hrsg.) (2021): Psychologische Diagnostik. Berlin, Heidelberg: Springer, 6., vollständig überarbeitete Auflage.

Schuhmacher, M. (o. J.): Diagnostik - Wie bereite ich das Kind auf den Test vor?

<https://www.diepraxis.de/content/diagnostik>

(Zugriff am 26.02.2023)

Spoerl, H. (2002): Die Feuerzangenbowle. Eine Lausbüberei in der Kleinstadt. Düsseldorf: Droste (es handelt sich um einen Text von 1933).

Stauche, H. (2008): Normwerte der Testdiagnostik - ihre Berechnungen und Umrechnungen

ineinander. Institut für Erziehungswissenschaft, Universität Jena. [https://www.db-](https://www.db-thueringen.de/servlets/MCRFileNodeServlet/dbt_derivate_00036055/normwerte.pdf)

[thueringen.de/servlets/MCRFileNodeServlet/dbt\\_derivate\\_00036055/normwerte.pdf](https://www.db-thueringen.de/servlets/MCRFileNodeServlet/dbt_derivate_00036055/normwerte.pdf) (Zugriff am 28.01.2020)

Trolldenier, H.-P. (o. J.): Diagnostische Methoden IV: Schulleistungstests

(Testbeispiele. Kennen lernen unterschiedlicher schulrelevanter Testtypen und ihrer Anwendungsbedingungen). Universität Würzburg.

[https://wuecampus.uni-](https://wuecampus.uni-wuerzburg.de/moodle/mod/book/tool/print/index.php?id=322322)

[wuerzburg.de/moodle/mod/book/tool/print/index.php?id=322322](https://wuecampus.uni-wuerzburg.de/moodle/mod/book/tool/print/index.php?id=322322)

(Zugriff am 07.06.2023)

Universität Zürich (2023): Methodenberatung. Deskriptive, univariate Analyse

(Verteilungen).

[https://www.methodenberatung.uzh.ch/de/datenanalyse\\_spss/deskuniv.html](https://www.methodenberatung.uzh.ch/de/datenanalyse_spss/deskuniv.html)

(Zugriff am 01.06.2023)

Wirtz, M., Nachtigall, C. (2008): Deskriptive Statistik. Statistische Methoden für Psychologen Teil 1. Weinheim, München: Juventa, 5., überarbeitete Auflage.

Wyschkon, A., Esser, G. (2015): Testleiterfehler und Beurteilung von Testnormen:

Empfehlungen für Testentwickler und -anwender. In: Esser, G., Hasselhorn, M.,

Schneider, W. (Hrsg.) (2015): Diagnostik im Vorschulalter. Göttingen, Bern, Wien, Paris, Oxford, Prag, Toronto, Boston, Amsterdam, Kopenhagen, Stockholm, Florenz, Helsinki: Hogrefe, 165-179.

Zimmermann, A. (2010): Schulnoten aus schulpsychologischer Sicht. Köln.

[https://www.yumpu.com/de/document/read/26101699/schulnoten-](https://www.yumpu.com/de/document/read/26101699/schulnoten-ausschulpsychologischer-sicht-es-schulpsychologie)

[ausschulpsychologischer-sicht-es-schulpsychologie](https://www.yumpu.com/de/document/read/26101699/schulnoten-ausschulpsychologischer-sicht-es-schulpsychologie)

(Zugriff am 12.06.2023)

Zimmermann, A. (2011): Thema „Intelligenz“: Definition, Entwicklung, Bedeutung und Messung.

Informationen und Hinweise für Lehrerinnen und Lehrer. Köln. [https://docplayer.org/20287148-](https://docplayer.org/20287148-Thema-intelligenz-definition-entwicklung-bedeutungund-messung-informationen-und-hinweise-fuer-lehrerinnen-und-lehrer-fassungaugust-2011.html)

[Thema-intelligenz-definition-entwicklung-bedeutungund-messung-informationen-und-hinweise-](https://docplayer.org/20287148-Thema-intelligenz-definition-entwicklung-bedeutungund-messung-informationen-und-hinweise-fuer-lehrerinnen-und-lehrer-fassungaugust-2011.html)

[fuer-lehrerinnen-und-lehrer-fassungaugust-2011.html](https://docplayer.org/20287148-Thema-intelligenz-definition-entwicklung-bedeutungund-messung-informationen-und-hinweise-fuer-lehrerinnen-und-lehrer-fassungaugust-2011.html)

(Zugriff am 12.06.2023)



---

### Über den Autor

Frank Herold ist Dipl.-Sozialpädagoge und Dipl.-Heilpädagoge und arbeitete bis zu seinem Ruhestand 2014 in der Beratungsstelle für Eltern, Jugendliche und Kinder der Stadt Hamm. Seine Approbation als Kinder- und Jugendlichenpsychotherapeut gab er 2016 zurück. Er ist verwitwet und hat zwei Söhne. Diesen Aufsatz widmet er seiner Frau Doris.

### Kontakt:

[frankherold@gmx.de](mailto:frankherold@gmx.de)

---

gew-ansbach.de – 29.10.2023 – S. 21

Was Sie schon immer wissen wollten über Entwicklungs-, Intelligenz- und Schultests einschließlich pädagogischer Beobachtungsbögen

---