

# Assessment Analytics

Ähnlich wie bei den Learning Analytics ist das Assessment Analytics eine Analyse von Prüfungsdaten. Alle während der Durchführung generierten Daten von den Prüflingen und Inhalten dienen dem Feedback an Studierende und Lehrende. Die Analyse soll Ausschluss darüber geben, wie die tatsächliche Leistung der Studierenden und die Qualität der Lehre inkl. Prüfung sind. Ebenfalls zeigen die Daten, wo ggf. nachgebessert werden kann.

Die Einschätzung und Analyse der Ergebnisse kann Aufschlüsse zu den unterschiedlichen Lerngruppen bzw. der Heterogenität allgemein der Studierenden ergeben, um die Lehre entsprechend anzupassen. Durch eine Optimierung der Lehre könnte die Abbrecherquote sinken und einzelne Studierende individuell gefördert werden.

## Interpretation der Kennwerte

Hier finden Sie die Analytic-Werte in der Aktivität Test in Moodle: <https://kb.el.uni-leipzig.de/books/moodle/page/test-erstellen#bkmrk-5.3-statistiklegen-s>.

### Leichtigkeitsindex

Der Leichtigkeitsindex gibt an, wie viele Prüflinge die Frage korrekt beantwortet haben. Bei den Werten 0 oder 0% hat nicht ein Prüfling die Frage korrekt beantwortet. Liegt der Index bei 0,5 bzw. 50% wurde die Aufgabe zur Hälfte richtig gelöst. Bei 1 oder 100% haben alle Teilnehmenden die Frage korrekt beantwortet. Idealerweise sollte der Index bei 30-70% liegen. Wichtig: Bei mehreren Versuchen hat das auch einen erheblichen Einfluss auf diesen Wert, da dann Fragen eher richtig beantwortet werden.

### Trennschärfeindex

Die Trennschärfe bietet eine Orientierung über die Leistung eines einzelnen Prüflings. Dieser Wert bildet den Zusammenhang der erreichten Punktzahl der Frage und der im gesamten Test ab. Als Richtwerte gelten Werte über 0,3 als gut, zwischen 0,3 und 0,2 als akzeptabel und Werte unter 0,2 als inakzeptabel. Liegt die Trennschärfe um 0, dann wurde die Aufgabe von guten wie von nicht guten Prüflingen gleich gut/schlecht beantwortet. Bei negativen Werten haben gute Prüflinge die Frage eher falsch beantwortet als nicht so gute Kandidaten. Damit die Aufgabe eine gute Aufgabe wird, sollte sie zwischen leistungsstarken und -schwächeren Prüflingen differenzieren können.

### Standardabweichung

Dieser Fragenwert gibt den durchschnittlichen Abstand aller erreichten Punkte zum Mittelwert der erreichten Punkte an. Eine geringe Standardabweichung bedeutet, dass die Testergebnisse der Teilnehmenden durchschnittlich sehr nahe bei dem Durchschnittswert liegen. Ein hoher Wert lässt dagegen auf eine breite Streuung der Ergebnisse um das Durchschnittsergebnis schließen. Daraus kann man schließen, ob es sich eher um eine heterogene oder homogene Punkteverteilung handelt.

Die Standardabweichung sollte idealerweise zwischen 12 und 18% liegen.

## Crombach oder Crombachs Alpha (Koeffizient interner Konsistenz)

Je näher der Wert für Crombach an 1(%) liegt, desto höher ist die Konsistenz des betrachteten Tests. Dies bedeutet, dass die Schwierigkeit der verschiedenen Fragen tendenziell ähnlich bzw. konsistent ist und es keine Fragen gibt, welche wesentlich besser oder schlechter ausfallen.

Ziel ist es also, Fragen zu erstellen, die ein ähnliches Fachwissen abfragen und eine ähnliche Schwierigkeit besitzen. Jedoch ist hier auf ein ausgewogenes Gleichgewicht zu achten, dass durchaus auch schwierigere Fragen im Test enthalten sein sollen, um die Leistung der Prüflinge besser differenzieren zu können. In diesem Fall ist ein eher niedriger Wert zu erwarten.

Optimalerweise sollte der Crombachs Alpha über 75% (bzw. 0,7) liegen. Die Spannbreite geht von minus unendlich bis 1/100%.

## Schiefe der Punkteverteilung

Sie zeigt an wie weit die Verteilung der erzielten Testergebnisse nach rechts oder nach links geneigt ist. Im Allgemeinen wird ein Wert bis -1,0 empfohlen. Ein noch kleinerer Wert kann auf eine mangelnde Unterscheidung bei Studierenden hindeuten, die überdurchschnittlich gut abschneiden. Ein Wert über 1 kann eine zu geringe Trennschärfe kennzeichnen.

## Bewertungsverteilungsgraph

Entspricht die Punkteverteilung einer Normalverteilung? Oder verläuft sie steiler oder breiter?

## Fehlerquotient

Der Fehlerquotient sagt aus, ob die Punkteunterschiede zwischen den Personen eher auf Zufall oder auf unterschiedlicher Vorbereitung basieren?

## Standardfehler

Dieser Wert wird vom Fehlerquotienten abgeleitet und ist ein Maß dafür, wie viel zufällige Abweichung in jeder einzelnen Testbewertung steckt. Wenn z.B. der Standardfehler 10% beträgt und eine Person im Test 60% erreicht hat, liegt der tatsächliche Wissenstand der Person zwischen 50% und 70%.

## Median

Der Median gibt an, welche Punktzahl liegt bei 50% der Punkteverteilung liegen würde.

## Beabsichtigte und effektive Gewichtung

Bei der beabsichtigten Gewichtung wird angezeigt, mit welcher Gewichtung die Frage in die Gesamtwertung eingehen sollte. Ein Beispiel wäre, wenn die Frage eine Bewertung von 3 Punkten hat, das diese dann 30 Prozent des ganzen Tests an Punkten ausmacht.

Die effektive Gewichtung gibt an, zu wie viel die erreichte Durchschnittspunktzahl aller Teilnehmenden bei der Frage tatsächlich zum Durchschnittsergebnis des Gesamttest beiträgt. Zur Überprüfung sollte die Frage gestellt werden, ob die effektive Gewichtung nahe der beabsichtigten Gewichtung liegt oder gar mit ihr übereinstimmt. Ist der Wert effektiven Gewichtung deutlich höher, als beabsichtigt, hat die Frage einen größeren Einfluss auf das Testergebnis, als beabsichtigt.

# Fragen optimieren

Die oben erklärten Werte und noch viele andere sollen Lehrende unterstützen, ihre Fragen zu optimieren. Die Anzahl schwerer und leichter Fragen kann überprüft werden. Besonders auch wenn mehrere Werte nicht im empfohlenen Bereich liegen, sollte ein Reviewprozess gestartet werden. Dabei kommentiert eine weitere Lehrkraft die Frage und Antwortmöglichkeiten und gibt Hinweise auf Verbesserungen. Eine Orientierung bietet unser **Selbstlernkurs zu kompetenzorientiert Prüfen**.

---

Version #10

Erstellt: 11 April 2024 11:29:04 von Steffi

Zuletzt aktualisiert: 19 November 2024 13:18:46 von Steffi