

## *Elektronisches Supplement 1. Exkurs zu Testlets in CATs und FITs.*

Es lassen sich aus unserer Sicht inter- und intraindividuelle Testlets unterscheiden. Der Textverständnisitempool enthält interindividuelle Testlets. Andere Textverständnisitems enthalten meist intraindividuelle Testlets.

Ein *intraindividuelles Testlet* besteht in der Textverständnisdiagnostik aus Aufgaben, die sich auf den gleichen Text beziehen und von denselben Schülerinnen und Schülern (SuS) bearbeitet werden. Sie werden diskutiert, weil sie die Annahme ‚lokaler stochastischer Unabhängigkeit (LSU)‘ vieler IRT Modelle verletzen können (Eckes, 2014). Die Verletzung der LSU kann sich unter anderem auf die Reliabilitätsschätzung auswirken. Das gilt für FITs und CATs gleichermaßen.

In einem CAT können intraindividuelle Testlets zusätzlich die Itemselektion einschränken, weil die Items eines Testlets von einem CAT in dem Verbund präsentiert werden müssen, in dem sie kalibriert wurden (Magis et al., 2017). Die Items sollten in dem Verbund präsentiert werden, in dem sie kalibriert wurden, weil die Fähigkeitsschätzung eines CATs auf den Itemparametern beruht, die in der Kalibrierung festgelegt wurden. Es ist plausibel, dass sich die Schwierigkeit eines Items verändert, wenn es in einem Fall das einzige Item und in einem anderen Fall z.B. das dritte Item ist, das zu einem Text bearbeitet wird.

Ein *interindividuelles Testlet* besteht aus Aufgaben, die sich auf den gleichen Text beziehen, aber nicht von den gleichen SuS bearbeitet werden. Interindividuelle Testlets verletzen nicht die LSU. Das lässt sich mathematisch und inhaltlich begründen. Mathematisch gesehen gibt es keine (tetrachorische) Korrelation zwischen den Items eines interindividuellen Testlets, weil sie nie von derselben Person beantwortet werden. Inhaltlich gesehen spielen die Ursachen, die für die Verletzung der LSU bei intraindividuellen Testlets angeführt werden, wie Vorwissen zu einem Text oder Folgeeffekte zwischen Aufgaben, keine Rolle (Eckes, 2014). Das gilt für FITs, die im Multi-Matrix-Design administriert werden, und CATs gleichermaßen.

In einem CAT können intraindividuelle Testlets zusätzlich die Itemselektion einschränken, weil die Bearbeitung eines Items zum Ausschluss der anderen Items des Testlets führen muss (Magis et al., 2017). Auch hier gilt, dass die Items eines CATs in dem Verbund präsentiert werden sollten, in dem sie kalibriert wurden. Für einen CAT bedeuten interindividuelle Testlets in einem Itempool, dass die maximale Anzahl der Items, die einer Person präsentiert werden kann, der Anzahl der Testlets entspricht und nicht mehr der Anzahl der Items. Solange die Anzahl der Items, die von einer Testperson tatsächlich in einem CAT bearbeitet wird, aber deutlich geringer ist als die Anzahl der Testlets im Itempool, sollte der Einfluss von intraindividuellen Testlets auf die Itemselektion gering sein.

## Referenzen

- Eckes, T. (2014). Lokale Abhängigkeit von Items im TestDaF-Leseverstehen. *Diagnostica*, 61, 93–106.
- Magis, D., Yan, D., & Von Davier, A. A. (2017). *Computerized adaptive and multistage testing with R: Using packages catr and mstr*. Springer. Berlin.