

Reproduzierbarkeitskoeffizienten (REP) überführt werden:

$$\text{REP} = 1 - \frac{\text{Anzahl der Fehler}}{\text{Anzahl der Befragten} \cdot \text{Anzahl der Items}}$$

Für das Beispiel mit 8 Personen und 7 Items erhält man

$$\text{REP} = 1 - \frac{4}{8 \cdot 7} = 0,93.$$

Dieser Wert liegt über 0,9 und würde damit Modellkonformität der Skala signalisieren.

Ein weiteres Maß zur Prüfung der Modellkonformität stellt **Loevens H-Koeffizient** dar, der z. B. bei Roskam (1996, S. 439) beschrieben wird.

Die hier diskutierte Skala verdeutlicht, wie stark sozialwissenschaftliche Messinstrumente von kulturellen und historischen Rahmenbedingungen geprägt sind. So gehen alle Skalenitems ganz selbstverständlich davon aus, dass Menschen heiraten und Biografien in eine Phase »vor der Ehe« und eine Phase »in der Ehe« zerfallen. Wer diese Vorstellung nicht teilt, für den sind die Testitems sinnlos.

Generell ist bei der Formulierung von Items darauf zu achten, dass sie keine impliziten Aussagen enthalten, die vom Probanden möglicherweise nicht geteilt werden und ihm somit keine Möglichkeit zum adäquaten Antworten lassen. Ein Ausweg aus diesem Problem ist die Verwendung von vorgeschalteten **Filterfragen**, die unterschiedliche Personengruppen identifizieren, denen dann jeweils nur die zur aktuellen Lebenssituation oder zu den individuellen Lebenseinstellungen passenden Fragen vorgelegt werden (► S. 244).

Edwards-Kilpatrick-Skala

Dieser von Edwards und Kilpatrick (1948) entwickelte Skalentyp vereinigt die von Thurstone, Likert und Guttman entwickelten Ansätze. Die Konstruktion beginnt mit der Sammlung eines Satzes dichotomer Items, der – wie bei der Thurstone-Skala – Experten mit der Bitte vorgelegt wird, die Intensität der mit der Bejahung (richtigen Lösung) eines Items zum Ausdruck gebrachten Merkmalsausprägung einzuschätzen. Es folgt die Aussortierung uneindeutig bewerteter Items. Die verbleibenden Items werden als Items mit vorgegebenen Antwortmöglichkeiten (6 Kategorien, die bei Einstellungsitems äquidistant gestufte Zustimmung repräsen-

tieren) einer für die Testanwendung repräsentativen »Eichstichprobe« zur Bearbeitung vorgelegt. Diese Itembeantwortungen liefern – wie bei der Likert-Skala – das Material für eine Trennschärfeanalyse, die zu einer weiteren Itemselektion führt. Von den trennscharfen Items werden schließlich nur diejenigen Items als dichotome Items zu einer Testskala vereinigt, die die Kriterien einer Guttman-Skala erfüllen.

Die Konstruktion dieser Skala ist damit sehr aufwendig und dürfte sich für eine einmalige Merkmalsmessung nur selten lohnen. Allerdings bietet sie eine gute Gewähr, dass tatsächlich eine Testskala mit überdurchschnittlichen Eigenschaften resultiert.

Rasch-Skala

Dieser Skalentyp, dessen theoretischer Hintergrund bereits auf ► S. 208 f. zusammengefasst wurde, basiert auf der Annahme, dass die Wahrscheinlichkeit der Lösung einer Aufgabe von der Ausprägung eines latenten Merkmals bei den untersuchten Personen abhängt (Personenparameter). Ausgehend von einem Satz inhaltlich homogener Items mit alternativen Antwortvorgaben, die als potenzielle Indikatoren des latenten Merkmals geeignet erscheinen, ermittelt man für jede Person die Anzahl gelöster Items. Es werden dann Personenparameter bestimmt, die die Wahrscheinlichkeit für das Zustandekommen der individuell erreichten Anzahl gelöster Aufgaben maximieren. Man nimmt hierbei an, dass die Wahrscheinlichkeit der Lösung eines Items ausschließlich von der Fähigkeit der Person und der Schwierigkeit des Items abhängt; die Art der Beantwortung eines Items ist also davon unabhängig, welche anderen Items die Person bereits bearbeitet hat (Prinzip der »lokalen stochastischen Unabhängigkeit«). Psychologisch gesehen bedeutet diese Forderung, dass die Itembeantwortungen von Übungs-, Ermüdungs- oder Positionseffekten unabhängig sind. Formal hat dieses Prinzip zur Konsequenz, dass sich die Wahrscheinlichkeit für die Gesamtanzahl gelöster Items für eine Person mit bestimmter Fähigkeit aus dem Produkt der Wahrscheinlichkeiten für die Lösung der einzelnen Items ergibt (genauer hierzu z. B. Amelang & Zielinski, 2002, Kap. 2.1.2.1; Rost, 2004, Kap. 2.3.4).

Die Schätzung der Itemparameter (Schwierigkeiten) erfolgt in ähnlicher Weise. Die Wahrscheinlichkeit, dass ein Item von einer bestimmten Anzahl von Personen