

# Itemanalyse ohne SPSS – alles auf einen Streich

DR. HELMUT STAUCHE UND CAND. PAED. NADINE WERLICH

Eine unabdingbare Prozedur bei der Konstruktion eines standardisierten Tests ist die Item- und Testanalyse. Unsere Erfahrungen besagen, dass sich viele Testautoren, die SPSS zur Güteüberprüfung eines neuen Tests einsetzen, auf einen der Analyseparameter, die innere Konsistenz aller eingebrachten Items<sup>1</sup> beschränken. Damit ist zwar eine wichtige Aussage zur Homogenität der getesteten Items gegeben, weitere unverzichtbare Aussagen bleiben so jedoch unbeachtet.

Nun ist es mit SPSS durchaus möglich, weitere Analyseparameter wie Schwierigkeitsindex und Trennschärfeindex zu gewinnen. Allerdings ist dies zumindest bezüglich des Schwierigkeitsindex nicht durch einen Mausklick zu erreichen, sondern setzt eine gewisse Mühe bei der Datenverarbeitung voraus.

Beide Parameter Schwierigkeitsindex und Trennschärfeindex gehören in die so genannte „klassische Testtheorie“, die vom differenzialpsychologischen Ansatz her entworfen wurde. Das heißt, der Test soll so gut wie möglich zwischen Testanden mit hoher und niedriger Merkmalsausprägung trennen, gleich ob es sich um eine Leistungs- oder um ein Persönlichkeitsmerkmal handelt.

## A. Der Schwierigkeitsindex

Unsinnig im oben erörterten Sinne wären Items eines Leistungstests, die von (fast) allen Testanden gelöst werden, ebenso wie diese, die von (fast) keinem Testanden gelöst werden. Übertragen auf Persönlichkeitsmerkmale heißt dies, dass Items (hier: dichotome<sup>2</sup> Items) wertlos sind, die von (fast) allen bejaht oder von (fast) allen verneint werden. Eine feine Differenzierung der Testanden nach dem betreffenden Merkmal, das heißt eine Einordnung des Testanden in die diesem Merkmal unterstellte Normalverteilung, ist so unmöglich.

Aus dieser Prämisse darf geschlossen werden, dass ein Test im optimalen Falle nur aus solchen Items bestehen sollte, die jeweils hälftig gelöst vs. nicht gelöst bzw. hälftig bejaht vs. verneint werden. Soweit ist das richtig, dennoch besteht die theoretische Möglichkeit, dass die Testandenmengen der Löser/Bejaher und Nichtlöser/Verneiner bei jedem Item dieselben sind. Ein solcher Ausgang des Tests wäre fatal, denn es gäbe keine Differenzie-

---

<sup>1</sup> meist als Cronbach's  $\alpha$  berechnet, das die Interkorrelation aller Items aufnimmt, vgl. S. 5

<sup>2</sup> das heißt *zweiwertige* Items, z. B. mit den Codes 0 und 1

rung sondern lediglich eine Zweierklassifizierung der Ergebniswerte. Nicht zuletzt ist diese Überlegung (neben der praktischen Erkenntnis, dass sich ein Test mit Items eines konstanten Schwierigkeitsindex von 0,5 sowieso nicht konstruieren lässt), dafür zuständig, dass Schwierigkeitsindizes in einem recht weiten Bereich als brauchbar angesehen werden. Im Allgemeinen werden mindestens 20 % Löser/Bejaher und höchstens 80 % Löser/Bejaher eines Items für „gut“ gehalten ( $0,2 \leq p \leq 0,8$ )<sup>3</sup>.

Der Schwierigkeitsindex eines Items wird also definiert als  $p = \frac{\text{erreichter Score}}{\text{erreichbarer Score}}$  und hat

damit den Definitionsbereich  $0 \leq p \leq 1$ , in der Praxis oft auch als  $0\% \leq p \leq 100\%$  angegeben.

Die Anwendung dieser Definition auf dichotome Items (richtig – falsch; ja – nein; trifft zu – trifft nicht zu; u.ä.) bedeutet, dass für jedes Item die Anzahl der Löser/Bejaher durch die Anzahl aller Mitglieder der Pilotstichprobe<sup>4</sup> zu dividieren ist.

Genauso gut kann diese Definition angewendet werden, wenn die die zur Reaktion auf die Items verwendete Likert-Skala nicht dichotom, sondern mehr als zweifach gestuft ist.

Beispiele hierfür sind:

- a. Die praktizierte Häufigkeit einer beliebigen Handlung, die mit mehr als zwei Werten von „nie“ bis „immer“ gestuft wird<sup>5</sup>;
- b. Ein Leistungstest mit komplexeren Aufgaben, die nicht nur dichotom (0 = nicht erfüllt vs. 1 = erfüllt) bewertet werden, sondern mit drei Codes: 0 = nicht erfüllt, 1 = teilweise erfüllt, 2 = erfüllt.

Allerdings ist zu beachten, dass der Schwierigkeitsindex nur dann richtig berechnet wird, wenn die kleinste Merkmalsausprägung mit der Null codiert wird. Für dichotome Items also 0 und 1, am Beispiel einer 4-fach-Stufung demnach 0 bis 3.

Der Schwierigkeitsindex bei beispielsweise 4-fach gestuften Items wird dann folgendermaßen berechnet:

Summe aller Codes 0 bis 3 über die Testanden der Pilotstichprobe hinweg (= erreichter Score) geteilt durch die dreifache Anzahl der Testanden der Pilotstichprobe (= erreichbarer Score: ist dann erfüllt, wenn alle Testanden den höchsten Wert – die 3 – angekreuzt hätten).

<sup>3</sup> p = Schwierigkeitsindex. p steht in statistischen Formeln fast ausschließlich für probability / Wahrscheinlichkeit. Dies trifft auch hier zu. Der in der Pilotstichprobe berechnete Wert des Schwierigkeitsindex drückt die Wahrscheinlichkeit aus, mit der auch die Grundgesamtheit dieses Item lösen/bejahen würde.

<sup>4</sup> Der erreichbare Score stellt sich dann ein, wenn alle Testanden das Item positiv beantworten, also  $n \cdot 1 = n$ .

<sup>5</sup> Sinnvoll ist in diesem Falle, auf dem Testbogen lediglich die Extreme zu attribuieren, oder selbst diese nicht verbal auszuführen sondern nur mit + und – zu kennzeichnen. So darf dem Testanden unterstellt werden, dass er die Abstände benachbarter Stufen als gleich groß auffasst und man (quasi)metrische Variablen erzielt.

In beiden Fällen – dichotome und nicht dichotome Items – kommt es im Falle der Messung eines Personmerkmals nicht darauf an, ob das Item „positiv gepolt“ ist (das heißt, bei Bejahung die stärkere Ausprägung des Merkmals ausdrückt) oder ob es „negativ gepolt“ ist. Viele Testautoren verwenden bewusst eine Mischung positiv und negativ gepolter Items, um den Test aufzulockern.

Diese simplen Bemerkungen zum Schwierigkeitsindex mögen hier genügen. Eine tiefgründigere Auseinandersetzung, die auf die Probleme dieser Definition verweist und Lösungsmöglichkeiten erörtert, findet sich u.a. bei Fisseni (1990, S. 30ff.)<sup>6</sup>.

## B. Der Trennschärfeindex

Die Trennschärfe gibt an, wie ähnlich die Items eines Tests von den Testanden der Pilotstichprobe gelöst/beantwortet werden – anders ausgedrückt, inwieweit die Menge der Löser/Bejaher über die gesamte Itemmenge hinweg konstant bleibt. Nach dieser Definition muss ein Assoziationskoeffizient zwischen dem einzelnen Item und dem Testscore berechnet werden. Wenn das einzelne Item mit der Summe aller, dem Testscore, einen starken Zusammenhang aufweist, fügt es sich inhaltlich gut in den Verbund der Items ein, es ist kein Quertreiber, kein Außenseiter. Es besitzt dann eine hohe Trennschärfe.

Sehr häufig wird – wenn (quasi)metrische<sup>7</sup> Variablengüte vorausgesetzt werden darf –, das einzelne Item mit dem Gesamtscore nach Pearson<sup>8</sup> korreliert. Für dichotome Verhältnisse genügt der so genannte punktbiserielle Korrelationskoeffizient, der die Mittelwerte zweier Klassen (hier des Gesamttests und dem Wert des Items) ins Kalkül nimmt.

Vergleichsrechnungen zeigen, dass auch die Anwendung des Pearson'schen Maßkorrelationskoeffizienten auf dichotom zu beantwortende Items zu ausreichend genauen Werten führen, so dass im weiter unten beschriebenen Programm nicht zwischen diesen Fällen unterschieden wird und immer Pearson's r berechnet wird.

Ein größerer Fehler entstünde, wenn man den gesamten Testscore zur Korrelation heranzöge, denn in ihm steckt ja auch das zu korrelierende Item. Der Trennschärfeindex wird so fälschlicherweise erhöht, die Verfälschung nimmt allerdings mit der Anzahl der Items ab.

Der Ausweg besteht darin, vom Testscore den Score des jeweils zu berechnenden Items zu subtrahieren und letzteren nur mit dem jeweiligen Restscore zu korrelieren. Dieses Verfahren nennt man Teil-Ganz-Korrektur.

<sup>6</sup> Fisseni, H.-J.: Lehrbuch der psychologischen Diagnostik. Hogrefe: Göttingen-Toronto-Zürich 1990.

<sup>7</sup> vgl. die einschlägige Fußnote auf S. 2

<sup>8</sup> Maßkorrelationskoeffizient nach Pearson: 
$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{s_x s_y (n-1)}$$

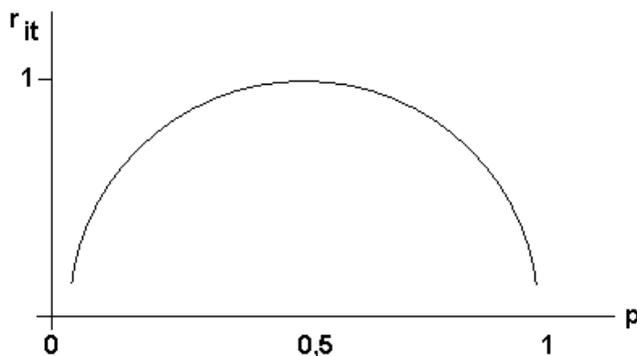
An vielen Stellen wird der Trennschärfeindex mit dem Formelzeichen  $r_{it}$  belegt ( $r$  steht für Korrelation,  $i$  für Item und  $t$  für Test).

Auch beim Trennschärfeindex soll auf das Optimum und den auf erwünschten Bereich des Parameters eingegangen werden: Optimal – und völlig illusorisch – wären Items, die alleamt mit dem Score der jeweils anderen zu  $r_{it} = 1$  korrelieren. Dies ist nur dann erfüllt, wenn die Testanden bei jedem Item dieselbe Antwortstruktur aufweisen. Dies zeigt, dass auch hier das Optimum lediglich ein formal-statistisches und kein inhaltliches ist, denn die Menge dieser Items würde nicht nur in zwei untereinander gleiche Klassen teilen wie beim Schwierigkeitsindex, sondern zu einer einzigen, untereinander völlig undifferenzierten Itemmenge führen. Im Allgemeinen gelten Items mit einem  $|r_{it}|$  von 0,3 und größer als „gut“<sup>9</sup>.

Anders als beim Schwierigkeitsindex kommt es beim Trennschärfeindex auf die einheitliche Polung aller Items an.

### C. Zusammenhang der Güteparameter

#### Zusammenhang $p$ und $r_{it}$ (Schwierigkeitsindex und Trennschärfeindex)



Dieser lässt sich durch ein auf den Kopf gestelltes U grafisch darstellen:

Je weiter der Schwierigkeitsindex vom Optimum 0,5 abweicht, desto geringer fällt im statistischen Mittel der Trennschärfeindex aus. Beide Optima fallen im Punkt ( $p = 0,5 \leftrightarrow r_{it} = 1$ ) zusammen.

#### Zusammenhang $r_{it}$ und $r_{ii}$ (Trennschärfeindex und Interkorrelation)

Beide Parameter stehen in einem positiven Zusammenhang. Hohe Interkorrelation<sup>10</sup> besagt, dass alle Items mehr oder weniger stark zu diesem hohen Wert beitragen. Damit erhöht sich aber auch die Wahrscheinlichkeit, dass das  $i$ -te Item mit seinem  $r_{it}$  hoch mit dem Gesamtscore korreliert.

<sup>9</sup> Zu beachten ist, dass ein Item – genau inverses Antwortverhalten bei umgekehrter Polung vorausgesetzt – zum selben Wert des Trennschärfekoeffizienten führt, allerdings mit umgekehrtem Vorzeichen. In unserem Programm besteht die Möglichkeit der Umkodierung der umgekehrt gepolten Items. Deshalb werden dort nur Trennschärfeindizes mit  $>+0,3$  (bzw. der individuell gewählten Grenze; vgl. S. 6) als gut anerkannt.

<sup>10</sup> Im Allgemeinen das arithmetische Mittel aus den  $n(n-1)/2$  einzelnen Korrelationskoeffizienten (jedes Item wird mit jedem korreliert.)

## D. Reliabilitätskoeffizienten

Wie bereits erwähnt, werden diese werden nach Erfahrung der Autoren von nicht wenigen Anwendern im Umgang mit Tests ausschließlich ins Kalkül genommen, ohne die vorab beschriebenen Indizes in Augenschein zu nehmen.

Bei ihrer Berechnung ist zwischen dichotomer und mehrfach skaliertes Itembeantwortung zu unterscheiden.

Besteht die Reaktion auf ein Testitem lediglich in „richtig – falsch“, „ja – nein“, „trifft zu – trifft nicht zu“ u.ä., dann ist Kuder-Richardson's Alpha für dichotome Items anzuwenden:

$$\alpha = \frac{n}{n-1} \cdot \left( 1 - \frac{\sum_{i=1}^k p_i \cdot (1-p_i)}{s_t^2} \right) \quad (\text{Kuder-Richardson's Alpha, KR-20 }^{11})$$

Ist dagegen die Reaktionsskala mindestens dreifach gestuft, dann ist Cronbach's Alpha der angemessene Koeffizient:

$$\alpha = \frac{n \cdot \bar{r}_{ii}}{1 + (n-1) \cdot \bar{r}_{ii}} \quad (\text{Cronbach's Alpha }^{12})$$

Alpha kann i.a.<sup>13</sup> in beiden Formeln Werte zwischen 0 und 1 annehmen. Je höher der Wert, desto besser ist die innere Konsistenz und damit auch die Reliabilität des Tests bzw. des Subtests.

## E. Die Exceldatei „itemparameter.xls“ mit ihrem VBA-Programm

Nach dem Öffnen der Datei sehen Sie zunächst das Blatt „start“ mit wenigen Bemerkungen zu Funktion und Vorgehen.

<sup>11</sup>  $n$  = Anzahl der Items,  $k$  = Anzahl der Fälle,  $p_i$  = Schwierigkeitsindex des  $i$ -ten Items,  $s_t^2$  = Gesamtvarianz aller Items des Tests, die sich bei bekannten Schwierigkeitsindizes nach

$s_t^2 = \sum_{i=1}^k p_i \cdot (1-p_i) + \sum_{i \neq j} (p_{ij} - p_i \cdot p_j)$  berechnet, worin  $p_{ij}$  den „Schwierigkeitsindex“ des Itempaars  $i$ - $j$

bedeutet. Quelle: Berres 1987, S. 44

<sup>12</sup>  $n$  = Anzahl der Items,  $\bar{r}_{ii}$  = arithmetisches Mittel aller bivariaten Korrelationen zwischen den Items.

Quelle: Ludwig-Mayerhofer 1999.

Die hier und in vielen weiteren Quellen übereinstimmend genannte Formel ist leider nicht die Berechnungsgrundlage für Cronbach's Alpha unter Reliability in SPSS. Dort wird  $\alpha = \frac{n \cdot \bar{r}_{ii}}{1 + n \cdot \bar{r}_{ii}}$  zugrunde gelegt und mithin

die auf S. 3 für den Trennschärfeindex erläuterte Teil-Ganz-Korrektur nicht berücksichtigt.

<sup>13</sup> Im besonderen Fall, wenn die mittlere Kovarianz zwischen den Items negativ ist, kann der Wert für Alpha auch negativ sein. Eine solche Skala/Subskala ist nicht brauchbar. Evtl. haben Sie den Fehler gemacht, ein oder mehrere negativ gepolte Items vor der Analyse nicht zu invertieren.

In jedem Falle ist vor einer neuen Berechnung der Button **AUSGANGSZUSTAND** zu bedienen<sup>14</sup>. Damit wird das Blatt „itemparameter“ in eben diesen Zustand versetzt und im Blatt „start“ auf die Standardkonstellation des kleinsten und größten Skalenwertes (Null und Eins) zurück gestellt sowie die o.g. Standards für den Schwierigkeits- und Trennschärfeindex eingestellt.

Als nächstes sind mit Hilfe der Drehfelder der Minimalwert und Maximalwert der Likert-Skala einzustellen. Das heißt, es sind der minimal bzw. maximal mögliche Wert zu nehmen, unabhängig davon, ob Randwerte evtl. von der gesamten Stichprobe überhaupt nicht angekreuzt wurden. Programmiert wurden Minimalwerte von –3 bis 1 sowie Maximalwerte von 1 bis 7. Damit sind auch nullsymmetrische Stufungen, z.B. –3 bis +3, möglich. Eine Überschneidung ist nur bei der 1 möglich, dieser Fehler wird jedoch vom Programm gemeldet.

Eine weitere Einstellmöglichkeit gibt es unter dem Punkt 3. im Blatt „start“. Dort lassen sich die Gütetoleranzen für beide Indizes in sinnvollen Grenzen um die Standards herum variieren. Für den Schwierigkeitsindex ist in 5%-Schritten eine Abstufung von 10%–90% bis hin zu 40%–60% vorgesehen, für den Trennschärfeindex in 0,05-Schritten zwischen 0,10 und 0,40.

Nach diesen ersten Schritten klickt man den einen der Buttons **ZUM BLATT ITEMPARAMETER** bzw. **ZUM BLATT DATENSATZ**.

Ersterer Button ist zu wählen, wenn Sie die Daten von Hand eintippen oder aus einem vorhandenen Datensatz kopieren wollen.

Im Blatt „itemparameter“ stehen Ihnen die weißen Zellen für das Einfügen des eigenen Datensatzes zu Verfügung. Alle anderen Zellen werden nur vom Programm beschrieben und sind nicht frei gegeben. Tabu sind also die Spalte A, sowie die Spalten IQ bis IV. In die Spalte A schreibt sich automatisch die laufende Nummer des Testanden und die Spalten

---

<sup>14</sup> Vor dem Bedienen der Buttons sind die VBA-Programme zu aktivieren. Dazu folgende Hinweise:

- bei MS Office 2003:  
Extras => Makro => Sicherheit bedienen und dort den Punkt bei Stufe 'Mittel' setzen. Nach erneutem Öffnen der Datei 'Makros aktivieren' wählen.
- bei MS Office 2007:  
Entwicklertools => Makrosicherheit und dort Punkt bei 'Alle Makros mit Benachrichtigung deaktivieren' setzen. Nach erneutem Öffnen der Datei auf Optionen neben der Sicherheitswarnung klicken und 'Diesen Inhalt aktivieren' wählen. Sollte in der Funktionsleiste 'Entwicklertools' nicht angezeigt werden, klicken Sie auf das Symbol links oben, danach unten rechts auf 'Excel-Optionen' und setzen den Haken bei 'Entwicklertools in der Multifunktionsleiste anzeigen'.
- bei MS Office 2010:  
Entwicklertools => Makrosicherheit und dort Punkt bei Alle Makros mit Benachrichtigung deaktivieren setzen. Nach erneutem Öffnen der Datei auf Optionen neben der Sicherheitswarnung klicken und Diesen Inhalt aktivieren wählen. Sollte in der Funktionsleiste Entwicklertools nicht angezeigt werden, auf Datei klicken, danach Optionen, Menüband anpassen und den Haken bei Entwicklertools (rechts) setzen (falls Entwicklertools noch nicht rechts steht, links auswählen und Hinzufügen).

IQ bis IV werden für diverse Zeilen- und Teilzeilensummen für die Berechnung des Trennschärfeindex benötigt.

In die freien Spalten können Sie also maximal 249 Items auf ein Mal eintragen, was wahrscheinlich für fast alle Fälle genügt.

Bezüglich des Umfangs der Testandenstichprobe ist quasi keine Obergrenze gesetzt. Ab Zeile 6 könnten Tausende Fälle eingetragen werden. Unsere Programmierung ist für maximal 4000 Fälle vorgesehen, um die Abarbeitung von Schleifen nicht unnütz zu verlängern.

Sie können den Datensatz direkt in das Blatt „itemparameter“ eintippen. Sofern er jedoch bereits in einer anderen Excel-Datei oder im SPSS-Datenblatt existiert, kann er kopiert werden. Wie geschieht das?

- an der Ausgangsstelle Datensatz ohne Variablenüberschriften markieren
- rechte Maustaste ⇒ Kopieren
- kontrollieren, dass im Blatt „itemparameter“ der Cursor in der Zelle B6 steht
- rechte Maustaste ⇒ Einfügen
- Endkontrolle, ob das Kopieren gelungen ist.

Anders als bei SPSS ist eine Definition von Fehlwerten (missing values) nicht vorgesehen. Hat ein Proband das eine oder andere Item nicht beantwortet, erscheint dies als leere Zelle.<sup>15</sup> Achten Sie auch darauf, dass keine „leeren“ Fälle, d.h. leere Zeilen, in Ihrem Datensatz sind. Dann würde die Aufnahme des Datensatzes für die Berechnung vor der Leerzeile unterbrochen.

Der andere Button auf dem Blatt „start“ führt Sie zum „datensatz“. Wenn Sie dorthin kopieren, dürfen in der Zeile 3 die Variablennamen (Itemnummern) stehen, dies ist aber nicht Bedingung. Der Vorzug dieses Blattes ist, dass Sie dorthin einen umfangreichen Datensatz kopieren können, auch dann, wenn nicht benötigte Variablen enthalten sind. Falls Ihre Items verschiedenen Subskalen angehören, kommt es nicht darauf an, vorher nach diesen zu sortieren. Alle Item-Spalten, die aus dem Blatt „datensatz“ in das Blatt „itemparameter“ übertragen werden sollen, müssen in der Zeile durch eine rote Zelle markiert werden. Dazu genügt ein Doppelklick mit der Maus in der betreffenden Zelle. Im Falle eines Irrtums

---

<sup>15</sup> Sollte in SPSS bereits ein Fehlwert definiert worden sein (oft als 9 oder 99), dann entfernen Sie diese Definition mit Hilfe vom SPSS wieder: Transformieren ⇒ Umkopieren in dieselben Variablen ⇒ alter Wert = [z.B.] 9 ⇒ neuer Wert = SYMIS (anstelle der [z.B.] 9 erscheint dann im Datensatz ein Punkt, der jedoch vom Programm entfernt wird). Denselben Effekt erreichen Sie auch, wenn Sie die Codierung der Fehlwerte zunächst mitkopieren und vor der Anwendung des Programms mit Hilfe von Suchen–Ersetzen den Fehlwertcode in eine Leerstelle verwandeln.

bewirkt ein wiederholter Doppelklick, dass die Zelle wieder ohne Farbfüllung erscheint. Auch für den Fall, dass alle belegten Spalten markiert werden sollen, gibt es einen Button.

Nunmehr darf auf den Button **1 . VORBEREITEN** geklickt werden. Damit wird eine ganze Serie vorbereitender Prozeduren vollzogen:

1. Sie werden an den Anfang geschickt, falls Sie bereits mit dem Programm arbeiten wollen, ohne vorher Daten eingegeben zu haben.
2. Sollten aus einem SPSS-Datensatz Punkte mit kopiert worden sein (die dort Systemfehlwerte darstellen), werden diese in Leerzellen verwandelt, damit das Programm richtig rechnen kann.
3. Das Programm erkennt, ob Ihr Datensatz so umzukodieren ist, dass danach sein Minimalwert = 0 ist (vgl. S. 2), und vollzieht dies automatisch. Damit Ihr ursprünglicher Datensatz nicht verschwindet (dies ist v.a. dann sinnvoll, wenn Sie die Daten erstmalig eingetippt haben), legt sich automatisch ein neues Blatt mit dem Namen „alter\_datensatz“ an, in das die Originaldaten geschrieben werden. Die Min-Max-Werte der Likert-Skala unter den Drehfeldern im Blatt „start“ werden ebenfalls mit umkodiert.

4. Sie werden nach evtl. zu invertierenden Items gefragt und tragen in Zeile 4 ein „u“ unter den Itemnummern ein, für die „negative Polung“ zutrifft.

Ermitteln Sie dazu anhand der Itemformulierungen (oder durch Befragen des Handbuchs, falls Sie die Güte eines auf dem Markt befindlichen Tests für Ihre eigene Stichprobe untersuchen wollen) die Nummern der negativ gepolten Items, d.h. derer, bei denen die Frage bzw. das Statement nicht die hohe sondern die niedrige Ausprägung des Merkmals ausdrückt.

5. Schließlich erhalten Sie die Aufforderung, die Zugehörigkeit zu maximal 5 Subskalen durch Schreiben der Ziffern 1 ...5 in die Zeile 5 einzutragen.

Diesen Arbeitsschritt lässt man weg, falls sich alle eingetragenen Items in einer Subskala befinden. In diesem Falle schreiben sich nur Einsen in die betreffenden Zellen der Zeile 5. Liegt aber eine Zuordnung der Items zu zwei oder mehr Subskalen vor, dann müssen alle Items auch zugeordnet werden.

Die Zuordnung darf völlig durcheinander sein, es muss also nicht bereits vorher durch Anwenden von Sortierfunktionen in Excel oder SPSS eine Ordnung nach Subskalen hergestellt werden.

Sind alle Items mit einer Subskalennummer versehen, ist die Vorbereitung abgeschlossen.

Mit dem Klicken des Buttons **2. INDIZES BERECHNEN** lösen Sie die Berechnung der Itemparameter aus  $p$  und  $r_{it}$  aus. Sollten Sie bis dahin formale Fehler gemacht haben, werden diese rechtzeitig vom Programm erkannt und Sie zur Korrektur aufgefordert. Nachstehende Fehler können auftreten:

1. Sie haben das Vorbereiten überhaupt noch nicht ausgeführt.
2. Sie haben Werte in Ihrem Datensatz, die außerhalb des im Blatt „start“ stehenden Intervalls liegen. Genannt wird Ihnen der erste gefundene falsche Wert. Entweder ist dieser Wert selbst falsch (dann im Datensatz korrigieren!) oder das Intervall wurde im Blatt „start“ falsch angegeben.

Eine Intervallkorrektur ist jetzt nur noch bezüglich des Maximalwertes möglich, denn der Minimalwert steht in jedem Falle bereits auf Null. Liegt der Fehler tiefer, müssten Sie von vorn anfangen und den Datensatz neu kopieren.

3. Sie haben im Falle der Subskalenzuordnung einem oder mehreren Items keine Subskala zugewiesen. Dies müssen Sie jetzt korrigieren.
4. Sie haben den Items mehr als 5 Subskalen zugewiesen. In diesem Falle werden Sie aufgefordert, den Datensatz so zu splitten, dass je Berechnung nicht mehr als 5 Subskalen in das Datenfeld aufgenommen werden. Dazu sind die Daten neu einzufügen. Es empfiehlt sich, die Möglichkeit der von-bis-Markierung im Blatt „datensatz“ dazu zu verwenden.
5. Es gibt den Fall, dass Testanden bei einem oder bei mehreren Items gleich antworteten. Dann hat das Item (bzw. haben die Items) die Varianz Null und die für den Trennschärfeindex und für Alpha nötigen Korrelationen können nicht berechnet werden.

In diesem Falle werden Sie gefragt, ob das Item ganz, d.h. für immer eliminiert werden darf. Falls Sie die Frage bejahen, wird dies im Blatt „itemparameter“ ausgeführt, im Blatt „datensatz“ bleibt es jedoch erhalten.

Ist alles richtig, erscheinen in den Zeilen 2 und 3 die auf zwei bzw. drei Dezimalstellen gerundeten Werte für den Schwierigkeitsindex<sup>16</sup> und den Trennschärfeindex.

Des Weiteren wird die Qualität der eingebrachten Items augenfällig gezeigt. Gemessen an den auf Seite 2 bzw. 4 erwähnten Standardgrenzwerten bzw. an den individuell gewählten Grenzwerten (vgl. S. 6) werden „gute“ Items grün und „schlechte“ rot markiert.

---

<sup>16</sup> Im Programm nicht im Wertebereich  $0 \leq p \leq 1$  angezeigt, sondern – wie in der Testdiagnostik verbreitet – als Prozentwert.

Der Button **3. ALPHA(S) BERECHNEN** liefert als letzte Programmierung dieser Arbeit den gleichnamigen Koeffizienten für die interne Konsistenz des Tests bzw. seiner einzelnen Subskalen. Das heißt, es entstehen automatisch so viele Alpha-Werte, wie es Subskalen gibt. Als Programmierbasis dienten uns die Formeln auf S. 5. Ob Kuder-Richardson's oder Cronbach's Alpha zu berechnen ist, entscheidet das Programm automatisch.

Ein zusätzliches Feature des Programms ist, dass man durch einen Doppelklick auf die Itemnummer (Zeile 1) oder auf die Zuordnungsnummer zur Subskala (Zeile 5) ein oder mehrere Items für die Berechnung des Alpha (bzw. der Alphas) eliminieren kann. Die Werte für den Schwierigkeitsindex und den Trennschärfeindex dieses/dieser Items erscheinen danach schwarz und der Button **3. ALPHA(S) BERECHNEN** kann erneut ausgeführt werden. Dieses Verfahren kann man beliebig oft wiederholen, um einen besseren Koeffizienten für die interne Konsistenz zu erzielen.

In erster Linie wird man das Verfahren dazu verwenden, das Alpha (bzw. die Alphas) für die um die „schlechten“ Items reduzierte Skala/reduzierten Subskalen zu ermitteln. Dabei müssen das Alpha (bzw. die Alphas) sich nicht unbedingt verbessern. Denn wurde ein Item nur wegen des zu schlechten Schwierigkeitsindex geschwärzt, kann dies zu einer Verschlechterung von Alpha führen.<sup>17</sup>

Sollte dieser Beitrag Ihr Interesse geweckt haben, dann nehmen Sie bitte Kontakt mit Dr. Helmut Stauche ([stauche46@gmail.com](mailto:stauche46@gmail.com)) auf.

## F. Literatur

- Berres, M. 1987: Stepwise Procedures for the Construction of Scales from Dichotome Items. In: Psychologische Beiträge, 29, 1987, S. 42-59
- Fisseni, H.-J. 1990: Lehrbuch der psychologischen Diagnostik. Verlag für Psychologie Hogrefe. Göttingen – Toronto – Zürich.
- Ludwig-Mayerhofer, W. 1999: Cronbachs Alpha. In: ILMES – Internet-Lexikon der Methoden der empirischen Sozialforschung. ([http://www.lrz-muenchen.de/~wlm/ilm\\_c4.htm](http://www.lrz-muenchen.de/~wlm/ilm_c4.htm) Last Update 30.12. 1999)

Der Anhang auf der nächsten Seite zeigt den Screenshot des oberen linken Ausschnittes des Blattes *itemparameter* nach vollständiger Berechnung.

---

<sup>17</sup> Beachten Sie, dass bei großen Datensätzen (viele Items und/oder viele Fälle) aufgrund der enormen Rechenaufgaben unter Umständen mehrere Minuten für die Berechnung der Parameter und des Alpha bzw. der Alphas benötigt werden.

Microsoft Excel - 0-bis-5er-Skala.xls

Frage hier eingeben

Item-Nummer

	A	B	C	D	E	F	G	H	I	J	K	
		1. Vorbereiten		2. Indizes berechnen			3. Alpha(s) berechnen					
1	Item-Nummer	1	2	3	4	5	6	7	8	9	10	
2	Schwierigkeitsindex	76,39	43,27	59,29	9,62	11,75	60,15	70,30	78,95	58,23	9,72	
3	Trennschärfeindex	0,301	0,291	0,389	0,193	0,167	0,360	0,180	0,394	-0,295	-0,162	
4	Alpha [Cronbach]	0,543	(SK 1)					0,478	(SK 2)			
5	Subskala-Nummer	1	1	1	1	1	1	2	2	2	2	
6	1	0	0	0	0	0	0	0	0	0	0	
7	2	5	0	5	0	0	2	5	5	0	0	
8	3	5	5	5	5	5	3	3	5	5	5	
9	4	5	0	0	0	0	4	5	4	5	0	
10	5	5	0	0	0	0	3	5	5	5	0	
11	6	5	0	5	5	5	4	5	5	0	0	
12	7	5	5	5	0	0	4	5	5	0	0	
13	8	0	0	5	0	0	3	2	4	5	0	
14	9	0	0	0	0	0	0	0	0	5	0	
15	10	5	0	0	0	0	2	1	4	0	0	
16	11	5	0	5	0	5	4	5	5	0	0	
17	12	5	0	5	0	0	4	4	5	5	0	
18	13	5	5	0	0	0	4	4	5	5	0	
19	14	5	0	0	0	0	3	5	5	0	0	
20	15	5	5	5	0	0	5	5	5	5	0	
21	16	5	5	5	0	0	2	4	5	0	0	
22	17	0	0	0	0	0	0	0	0	0	0	
23	18	5	5	5	0	0	4	5	5	5	0	
24	19	5	0	0	0	0	5	5	5	5	5	
25	20	5	0	0	0	0	4	5	5	5	0	
26	21	5	5	5	0	0	4	5	5	0	0	
27	22	5	5	5	5	0	3	5	5	5	5	
28	23	5	0	5	0	0	2	4	5	0	0	

start | itemparemeter | datensatz

Bereit

NF

Start | Posteingang - Mozilla Th... | 0-bis-5er-Skala.xls | Microsoft Visual Basic - 0... | itemparemeter.doc - Micr... | 22 | 13:10