

Beispiel zur Reliabilitätsanalyse einer Likertskala

Wir erläutern die Reliabilitätsanalyse einer Skala exemplarisch an einer Likertskala¹ zur Messung des **allgemeinen Selbstwertes**. Diese Skala entstammt der Konsumsuchtstudie von Gerhard Scherhorn. Der allgemeine Selbstwert umfaßt allgemeine Einstellungen (Attitüden) eines Individuums gegenüber sich selbst - sehr allgemein ausgedrückt: das „Selbstbewußtsein“ einer Person (vgl. Scherhorn et al 1993, S. 28ff)². Den Befragten wurden 10 Items vorgelegt, zu denen sie auf einer 6-stufigen Ratingskala Zustimmung oder Ablehnung äußern konnten:

Item	Wortlaut	Nr. im Datensatz
Item 1	Manchmal glaube ich, daß ich zu nichts gut bin	V226
Item 2	Ich finde mich ganz in Ordnung	V227
Item 3	Ich bin ein Niemand	V228
Item 4	Eigentlich bin ich mit mir ganz zufrieden	V229
Item 5	Manchmal wünschte ich, ich wäre nicht geboren	V230
Item 6	Ich wollte, ich könnte mehr Achtung vor mir haben	V231
Item 7	Manchmal fühle ich mich zu nichts nütze	V232
Item 8	Wenn ich mich mit anderen Menschen meines Alters vergleiche, schneide ich eigentlich ganz gut ab	V233
Item 9	Ich verachte mich	V234
Item 10	Ich bin zufrieden mit mir	V235

Ein hohes Rating der Items signalisiert Zustimmung. Die Items 1,3,5,6,7 und 9 sind also im Sinne der Messung von Selbstwert negativ gepolt. Ein hoher Wert (Zustimmung!) zeigt bei diesen Items einen geringen Selbstwert des/der Befragten an.

1. Transformation der Items

Zunächst werden die Items des Originaldatensatzes (**v226 bis v235**) in neue Variablen transformiert (**sw1 bis sw10**). Hierbei werden die negativ formulierten Items umgepolt, damit eine höhere Ausprägung jedes Items immer auch einen höheren Selbstwert anzeigt.³ Die Items haben alle die Ausprägungen 1 (lehne völlig ab) bis 6 (stimme völlig zu). Damit also bei den gedrehten Items die Ausprägung 6 zur 1 wird, subtrahieren wir die entsprechenden Items jeweils von 7:

¹ vgl. Kapitel 5.3.1 im Skript zur Vorlesung „Methoden der empirischen Wirtschafts- und Sozialforschung“, 2. Auflage.

² Scherhorn, Gerhard; Ariane Gläser; Michael Neuner; Gerhard Raab; Lucia Reisch (1993): *Indikatoren der Bevölkerungsumfrage*, Lehrstuhl für Konsumtheorie und Verbraucherpolitik Universität Hohenheim, Arbeitspapier 62, Stuttgart.

³ Diese Transformation müßte eigentlich nur für die „gedrehten“ Items durchgeführt werden. Es erhöht jedoch die Übersichtlichkeit während der Datenanalyse, wenn alle Items einer Skala einheitlich bezeichnet sind.

```
compute sw1 = 7-v226.  
compute sw2 = v227.  
compute sw3 = 7-v228.  
compute sw4 = v229.  
compute sw5 = 7-v230.  
compute sw6 = 7-v231.  
compute sw7 = 7-v232.  
compute sw8 = v233.  
compute sw9 = 7-v234.  
compute sw10 = v235.
```

Wir haben nun 10 neue Variablen (**sw1** bis **sw10**), jeweils mit den Ausprägungen 1 bis 6, wobei die Ausprägung 1 einen niedrigen Selbstwert und die Ausprägung 6 einen hohen Selbstwert anzeigt.

2. Überprüfung der Schwierigkeit

Um die Schwierigkeit der Items zu bestimmen, gibt es keine „Patentlösung“. Ursprünglich gibt die Schwierigkeit den Prozentsatz der Zustimmung zu einem Item an. Eine Strategie zur Bestimmung der Schwierigkeit könnte also die Aufteilung jedes Items in zwei Hälften sein (Dichotomisierung: Ausprägung 1 bis 3 entspräche Ablehnung, Ausprägung 4 bis 6 entspräche Zustimmung), um dann die prozentuale Häufigkeit für jede Hälfte abzulesen.

Eine andere Möglichkeit besteht darin, die Items auf einen Bereich von 0 bis 1 zu transformieren und den Mittelwert dieser normierten Items als Maß für die Schwierigkeit zu interpretieren. Diese Methode wurde im folgenden angewandt:

Zunächst wurden die Variablen auf den Bereich von 0 bis 1 normiert. Jede Variable hat die Ausprägungen von 1 bis 6. Durch Subtraktion von 1 ergibt sich ein neuer Wertebereich von 0 bis 5. Teilt man nun noch durch 5, so erhält man 6 Ausprägungen von 0 bis 1. Die neuen Variablen sind **schw1** bis **schw10**.

```
compute schw1 = (sw1-1)/5.  
compute schw2 = (sw2-1)/5.  
compute schw3 = (sw3-1)/5.  
compute schw4 = (sw4-1)/5.  
compute schw5 = (sw5-1)/5.  
compute schw6 = (sw6-1)/5.  
compute schw7 = (sw7-1)/5.  
compute schw8 = (sw8-1)/5.  
compute schw9 = (sw9-1)/5.  
compute schw10 = (sw10-1)/5.
```

Nun können wir uns die Mittelwerte dieser normierten Items berechnen lassen. Der Befehl dazu lautet:

```
DESCRIPTIVES  
VARIABLES=schw1 schw2 schw3 schw4 schw5 schw6 schw7 schw8 schw9 schw10  
/FORMAT=LABELS NOINDEX  
/STATISTICS=MEAN STDDEV MIN MAX .
```

Wir erhalten folgenden Output:

Number of valid observations (listwise) = 1456,00

Variable	Mean	Std Dev	Minimum	Maximum	Valid N	Label
SCHW1	,84	,24	,00	1,0	1511	
SCHW2	,76	,22	,00	1,0	1508	
SCHW3	,91	,20	,00	1,0	1506	
SCHW4	,77	,21	,00	1,0	1519	
SCHW5	,90	,21	,00	1,0	1508	
SCHW6	,83	,24	,00	1,0	1502	
SCHW7	,84	,25	,00	1,0	1504	
SCHW8	,73	,22	,00	1,0	1506	
SCHW9	,94	,16	,00	1,0	1506	
SCHW10	,77	,22	,00	1,0	1517	

In der Spalte **Mean** stehen die Mittelwerte der normierten Items, diese Mittelwerte können als Maß für die Schwierigkeit der Items (Prozentsatz der Zustimmung) interpretiert werden. Wir sehen, daß alle Items einen relativ hohen Anteil an Zustimmung haben (was generell nicht für eine besonders gelungene Auswahl der Items spricht)⁴. Insbesondere die Items 3, 5 und 9 haben eine besonders hohe Schwierigkeit, werden also von über 90% der Befragten zustimmend beantwortet. Es wäre also potentiell zu empfehlen, diese Items aus der Skala zu eliminieren. Betrachten wir nocheinmal die Formulierung der Items:

- Item 3: „Ich bin ein Niemand“
- Item 5: „Manchmal wünschte ich, ich wäre nicht geboren“
- Item 9: „Ich verachte mich“

Alle drei Items sind negativ formuliert (wurden also von uns bereits recodiert; s.o!). Zustimmung im Sinne der Selbstwertkala bedeutet also hier die Ablehnung der Items durch die Befragten. Angesichts der extremen Formulierung der Items ist es keine große Überraschung, daß über 90% der Befragten diese Items abgelehnt haben. Wir werden diese Items bei der endgültigen Skala deshalb nicht mehr verwenden.

3. Die Trennschärfe der Items

Um die Trennschärfe der Items zu berechnen, gibt es zwei Alternativen. Die eine Möglichkeit ist der t-Test auf Mittelwertunterschiede für jedes Item, getrennt nach den 25% der Befragten mit den höchsten und den niedrigsten Skalenwerten (der Rohskala!).⁵ Wesentlich besser, insbesondere bei großen Stichproben (hier z.B. $N > 1400$), ist die Berechnung der Trennschärfe als Korrelation des jeweiligen Items mit der Summe der übrigen Items der Skala. Dieses Maß wird im Rahmen der SPSS-Prozedur

⁴ Die verwendete Skala ist eine Subskala einer Selbstkonzeptskala und findet normalerweise eher im klinischen Bereich Verwendung. Dort ist es möglicherweise das Ziel, schwere Störungen der Persönlichkeit zu messen - diese sollten naturgemäß nicht bei der Masse der Bevölkerung vorliegen.

⁵ Dieses Verfahren wird in unserer Beispieldatei zum t-Test genauer erläutert (siehe T-TEST.DOC).

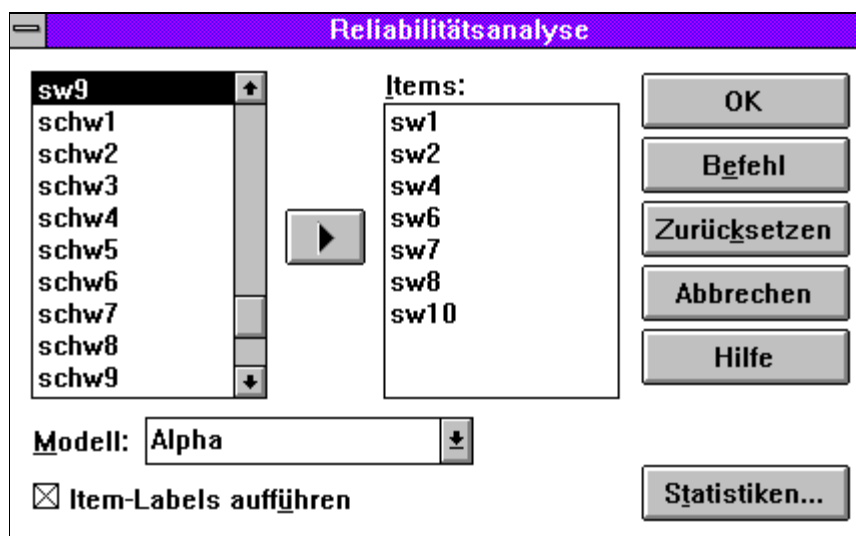
RELIABILITY als sog. **Corrected Item-Total-Correlation** berechnet. Grundsätzlich sollten alle Items einer Skala untereinander positiv korreliert sein. Zur Kontrolle sollte deshalb bei der **RELIABILITY**-Prozedur immer auch die Korrelationsmatrix der Items angefordert werden. Negative Korrelationen können zwei Ursachen haben:

- Entweder das Item ist tatsächlich ungeeignet für die Skala, in diesem Fall ist die Korrelation i.d.R. nahe Null,
- oder das Item wurde nicht (oder falsch) recodiert. Sollte dies der Fall sein, so müsste das Item negativ, aber hoch mit den übrigen Items korreliert sein.

Der Befehl zur Reliabilitätsanalyse⁶ der von uns untersuchten Skala lautet:

```
RELIABILITY  
/VARIABLES=sw1 sw2 sw4 sw6 sw7 sw8 sw10  
/FORMAT=LABELS  
/SCALE(ALPHA)=ALL/MODEL=ALPHA  
/STATISTICS=DESCRIPTIVE CORR  
/SUMMARY=TOTAL CORR .
```

Über das Menü in SPSS für Windows (Statistiken → Skalierung (MDS) → Reliabilität) gelangt man zu folgendem Dialogfenster:



Hier werden die Variablen der Skala in das Fenster „Items:“ übertragen.

Als „Modell“ wurde „Alpha“ ausgewählt - es wird also Cronbachs α berechnet.

Über den Button „Statistiken ...“ gelangt man zu folgendem Fenster:

Über den Button „Statistiken“ gelangt man zu folgendem Fenster:

⁶ Zur Reliabilität vgl. Kapitel 5.4.1.1 im Skript zur Vorlesung „Methoden der empirischen Wirtschafts- und Sozialforschung“, 2. Auflage.

Reliabilitätsanalyse: Statistiken

<p>Deskriptive Statistiken für</p> <p><input checked="" type="checkbox"/> Item</p> <p><input type="checkbox"/> Skala</p> <p><input checked="" type="checkbox"/> Skala, wenn Item gelöscht</p>	<p>Zwischen Items</p> <p><input checked="" type="checkbox"/> Korrelationen</p> <p><input type="checkbox"/> Kovarianzen</p>
<p>Zusammenfassung</p> <p><input type="checkbox"/> Mittelwert</p> <p><input type="checkbox"/> Varianzen</p> <p><input type="checkbox"/> Kovarianzen</p> <p><input checked="" type="checkbox"/> Korrelationen</p>	<p>ANOVA-Tabelle</p> <p><input checked="" type="radio"/> Keine</p> <p><input type="radio"/> F-Test</p> <p><input type="radio"/> Friedman Chi-Quadrat</p> <p><input type="radio"/> Cochran Chi-Quadrat</p>
<p><input type="checkbox"/> Hotellings T-Quadrat</p>	<p><input type="checkbox"/> Tukey Additivitätstest</p>

Wichtig ist hier vor allem die Aktivierung der Felder „Item“ und „Skala, wenn Item gelöscht“. Dadurch werden unter anderem die uns interessierenden Trennschärfekoeffizienten angefordert.

Ferner wird die Korrelationsmatrix der Items („Korrelationen“) und eine „Zusammenfassung“ dieser Matrix angefordert.

Wir erinnern uns, daß wir bereits 3 Items aufgrund ihrer Schwierigkeit aus der Analyse ausgeschlossen haben. Diese Items tauchen deshalb in der Itemliste nicht mehr auf. Wir erhalten nach Ausführung dieses Befehls folgenden (gekürzten) Output:

R E L I A B I L I T Y A N A L Y S I S - S C A L E (A L P H A)

		Mean	Std Dev	Cases
1.	SW1	5,1992	1,2208	1471,0
2.	SW2	4,7852	1,0961	1471,0
3.	SW4	4,8294	1,0691	1471,0
4.	SW6	5,1502	1,1895	1471,0
5.	SW7	5,2087	1,2192	1471,0
6.	SW8	4,6356	1,0876	1471,0
7.	SW10	4,8742	1,0791	1471,0

Zunächst werden nocheinmal die Mittelwerte und die Standardabweichung der Items angezeigt. Es folgt die Korrelationsmatrix:

Correlation Matrix

	SW1	SW2	SW4	SW6	SW7	SW8
SW1	1,0000					
SW2	,3543	1,0000				
SW4	,3117	,4999	1,0000			
SW6	,4666	,2977	,2946	1,0000		
SW7	,5562	,3161	,4042	,5427	1,0000	
SW8	,2740	,4428	,4309	,2711	,2944	
SW10	,2834	,4286	,5445	,2628	,3209	,4467

N of Cases = 1471,0

Wir sehen, daß keines der Items mit einem anderen Item negativ korreliert ist. Einen komprimierten Überblick über die Korrelationsmatrix liefert die folgende Tabelle:

Inter-item Correlations	Mean	M i n i m u m	M a x i m u m	R a n g e	M a x / M i n	V a r i a n c e
	,3831	,2628	,5562	,2934	2,1167	,0098

Die durchschnittliche Inter-Item-Korrelation (der Mittelwert der Korrelationsmatrix) liegt bei ,3831. Die kleinste Korrelation beträgt ,2628, die höchste Korrelation immerhin ,5562.

Es folgt die für uns interessanteste Tabelle:

Item-total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Alpha if Item Deleted
SW1	29,4833	22,1397	,5438	,3750	,7885
SW2	29,8973	22,8650	,5543	,3557	,7864
SW4	29,8532	22,6873	,5937	,4282	,7800
SW6	29,5323	22,6314	,5152	,3462	,7934
SW7	29,4738	21,6372	,5954	,4495	,7787
SW8	30,0469	23,3781	,5055	,3046	,7945
SW10	29,8083	23,1278	,5380	,3711	,7891

In den ersten beiden Spalten ist angegeben, wie sich der Mittelwert und die Varianz der Skala (also der Summe dieser Items) verändern würde, wenn das entsprechende Item nicht in die Skala mit aufgenommen würde. In der dritten Spalte steht die Trennschärfe jedes Items. Dies ist die Korrelation des Items mit der Summe der übrigen Items. Wir erinnern uns, daß dieser Koeffizient für die Trennschärfe eines Items wenigstens 0,5 betragen sollte (Daumenregel!). Dieses Kriterium wird von allen Items der Skala erfüllt. In der darauffolgenden Spalte steht die quadrierte multiple Korrelation jedes Items. Diese kann auch als Anteil der durch die anderen Items erklärten Varianz des Items interpretiert werden - also als ein angenäherter Wert für die Reliabilität des einzelnen Items. In der letzten Spalte wird angegeben, welches α die Skala erzielen würde, wenn das Item aus der Skala entfernt würde. Um beurteilen zu können, ob die Elimination eines Items in diesem Sinne eine Verbesserung oder Verschlechterung darstellen würde, muß der letzte Teil des Outputs beachtet werden:

Reliability Coefficients 7 items

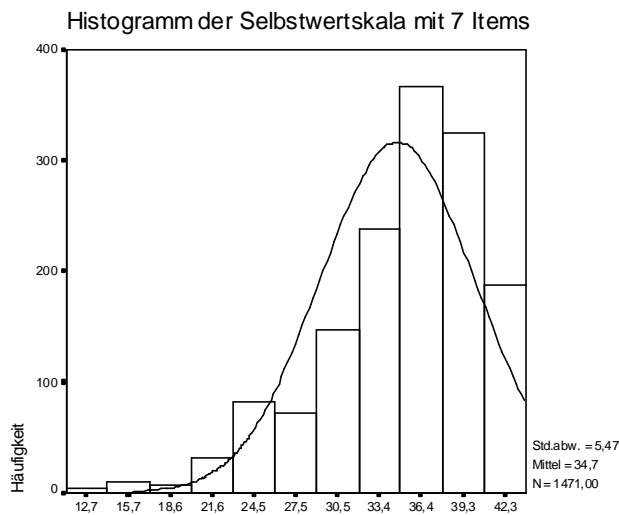
Alpha = ,8120 Standardized item alpha = ,8130

Dieses α ist die von der vorliegenden Skala erzielte Reliabilität (Cronbachs α). Wie schon angesichts der hohen Trennschärfen der Items zu vermuten war, läßt sich die Reliabilität der Skala nicht mehr steigern, wenn ein weiteres Item ausgeschlossen wird. Die Angabe **Standardized item alpha** gibt an, wie hoch α würde, wenn die Items vor dem Aufsumieren standardisiert würden. Durch die Standardisierung erhalten die Items die gleiche Varianz ($s^2 = 1$). Es kann dann die vereinfachte Formel für Cronbachs α verwendet werden.

Als letzter Schritt muß nun die endgültige Skala gebildet werden. Wir sumieren nun also die 7 Items zu einer Gesamtskala (**selwert**) auf:

`compute selwert = sw1+sw2+sw4+sw6+sw7+sw8+sw10.`

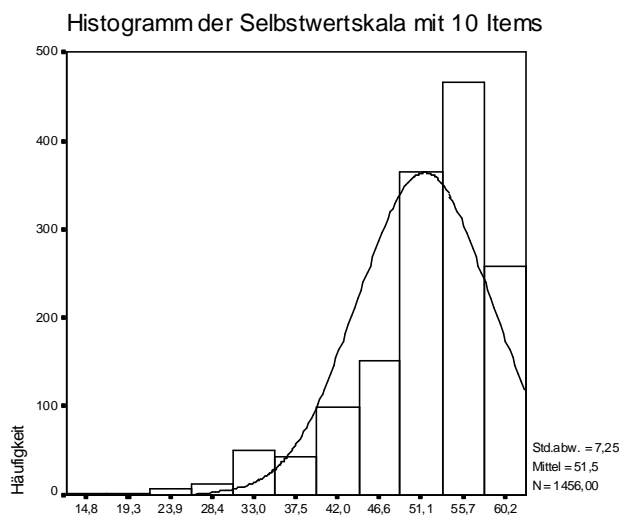
Um das Ergebnis zu kontrollieren, betrachten wir zum Abschluß die Verteilung der neuen Skala zur Messung des Selbstwertes. Dazu haben wir ein Histogramm mit Normalverteilungskurve angefordert (z.B. über die Prozedur **FREQUENCIES**):



Wir sehen, daß die Skala stark linksschief verteilt ist, was angesichts der hohen Werte für die Schwierigkeit der Items von vorneherein zu erwarten war. Immerhin ist die Skala eingipflig. Nur zu Demonstrationszwecken haben wir eine zweite Skala aus allen 10 Items gebildet:

`compute selwert1= sw1+sw2+sw3+sw4+sw5+sw6+sw7+sw8+sw9+sw10.`

Die Verteilung dieser Skala ist noch extremer linksschief (da die drei Items mit besonders hoher Schwierigkeit hier ebenfalls enthalten sind):



Die Verteilung weicht noch stärker von der theoretisch zu erwartenden Normalverteilung ab, als die aus 7 Items gebildete Skala. Der Ausschluß der 3 Items mit zu hoher Schwierigkeit hat sich also günstig auf die Verteilung der endgültigen Skala ausgewirkt.

4. Einige weitere Beispiele zur Beurteilung der Trennschärfe

Die Trennschärfe aller Items im ersten Beispiel war ausreichend hoch ($> .5$). Dies ist jedoch häufig nicht der Fall. Deshalb folgen nun einige kürzere Beispiele anderer möglicher Konstellationen, in denen die Trennschärfekoeffizienten nicht so günstig ausfallen:⁷

4.1 Ein Item hat eine wesentlich geringere Trennschärfe

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Alpha if Item Deleted
BSP1	-----	-----	, 5438	-----	, 7213
BSP2	-----	-----	, 5543	-----	, 7102
BSP3	-----	-----	, 3678	-----	, 7934
BSP4	-----	-----	, 5152	-----	, 7342
BSP5	-----	-----	, 5954	-----	, 7003
BSP6	-----	-----	, 5055	-----	, 7379
BSP7	-----	-----	, 5380	-----	, 7284

Reliability Coefficients 7 items

Alpha = , 7580

In diesem Fall hat das Item BSP3 eine viel zu niedrige Trennschärfe. Die Elimination dieses Items aus der Skala würde Cronbachs α von .7580 auf .7934 erhöhen. Die endgültige Skala würde also ohne dieses Item gebildet.

4.2 Viele Items haben eine zu geringe Trennschärfe

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Alpha if Item Deleted
BSP1	-----	-----	, 4438	-----	, 6073
BSP2	-----	-----	, 4543	-----	, 6052
BSP3	-----	-----	, 4678	-----	, 6034
BSP4	-----	-----	, 4552	-----	, 6054

Reliability Coefficients 4 items

Alpha = , 6280

Alle Items haben eine Trennschärfe $< .5$. Allerdings läßt sich durch eine Elimination irgendeines Items keine Verbesserung der Reliabilität der Gesamtskala erreichen, da die Skala aus zu wenig Items besteht (\rightarrow Reliabilitätsverbesserung durch Testverlängerung!). Entweder, man gibt sich an dieser Stelle mit dieser nicht sehr reliablen Skala zufrieden (was mangels Alternative häufig der Fall sein wird), oder man entscheidet, daß eine reliable Messung des theoretischen Konstruktes mit dieser Skala (genauer: mit diesen Items) nicht möglich ist.

⁷ Es handelt sich hier um rein hypothetische Beispiele, die nur der Erläuterung dienen. **Sämtliche Werte sind frei erfunden, ein Bezug auf diese Werte (z.B. in Hausarbeiten ...) ist deshalb nicht zulässig.** Die in diesem Fall weniger interessierenden Angaben des Outputs wurden weggelassen.

4.3 Zwei Items haben eine zu geringe Trennschärfe

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item- Total Correlation	Squared Multiple Correlation	Alpha if Item Deleted
BSP1	-----	-----	, 5438	-----	, 7373
BSP2	-----	-----	, 5543	-----	, 7252
BSP3	-----	-----	, 4678	-----	, 7404
BSP4	-----	-----	, 5552	-----	, 7154
BSP5	-----	-----	, 5438	-----	, 7373
BSP6	-----	-----	, 5643	-----	, 7152
BSP7	-----	-----	, 4573	-----	, 7434

Reliability Coefficients 7 items

Alpha = , 7480

Die Items BSP3 und BSP7 haben eine zu geringe Trennschärfe. Die letzte Spalte zeigt jedoch an, daß sich die Reliabilität der Skala durch den Ausschluß eines Items nicht erhöhen würde. Achtung: Hier besteht die Gefahr eines Trugschlusses! Es kann nämlich sein, daß die Reliabilität der Skala nach Ausschluß beider Items höher wäre. In diesem Fall ist es ratsam, eine weitere Reliabilitätsanalyse nach Ausschluß beider Items durchzuführen und deren Ergebnisse mit dem obigen Output zu vergleichen.