

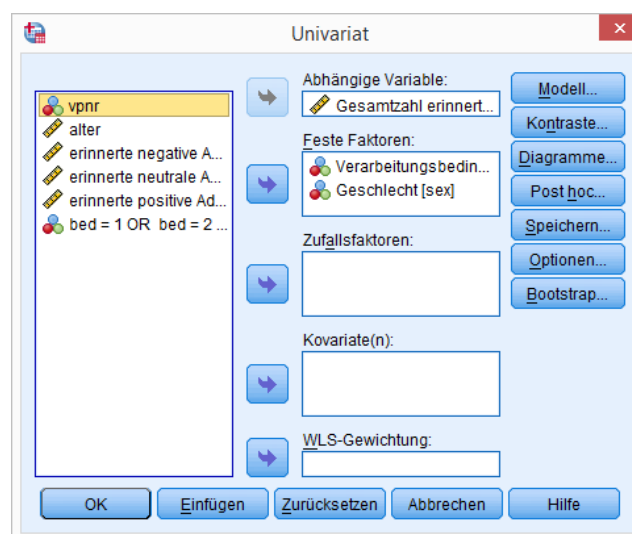
## Kapitel 6: Zweifaktorielle Varianzanalyse

Durchführung einer zweifaktoriellen Varianzanalyse ohne Messwiederholung	1
Effektstärke und empirische Teststärke einer zweifaktoriellen Varianzanalyse ohne Messwiederholung	8
Literatur	9

### Durchführung einer zweifaktoriellen Varianzanalyse ohne Messwiederholung

Dieser Abschnitt zeigt die Durchführung der in Kapitel 6 behandelten zweifaktoriellen Varianzanalyse ohne Messwiederholung mit SPSS. Als Beispiel dient wie schon in Kapitel 5 die Frage, ob es einen signifikanten Unterschied zwischen der Erinnerungsleistung von Adjektiven bei „struktureller“, „bildhafter“ oder „emotionaler“ Verarbeitung der Wörter gibt. Neben dem Faktor „Verarbeitungsbedingung“ soll hier nun zusätzlich der Einfluss des Faktors „Geschlecht“ auf die Erinnerungsleistung untersucht werden. Zusätzlich erlaubt die zweifaktorielle Varianzanalyse Aussagen über eine mögliche Wechselwirkung zwischen den beiden Faktoren.

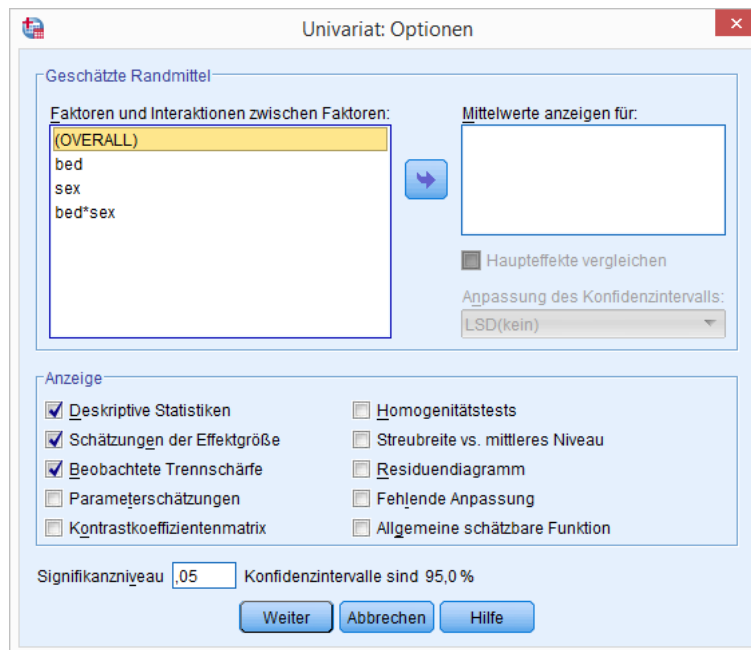
Gehen Sie in SPSS zu „Analysieren“ → „Allgemeines lineares Modell“ → „Univariat“. Dort können Sie die abhängige Variable auswählen, die Sie untersuchen möchten („Gesamtzahl erinnertes Adjektive“). Außerdem müssen Sie spezifizieren, welche Gruppen oder experimentelle Bedingungen das Verfahren miteinander vergleichen soll. Dazu bewegen Sie die Variable „Verarbeitungsbedingung“ in das Feld „Feste Faktoren“. Bewegen Sie nun zusätzlich die Variable „Geschlecht“ in das Feld „Feste Faktoren“. Sie untersuchen nun den Einfluss von zwei Faktoren auf die abhängige Variable, berechnen also eine zweifaktorielle Varianzanalyse. Durch Hinzufügen weiterer Faktoren könnten Sie beliebige mehrfaktorielle Varianzanalysen rechnen.



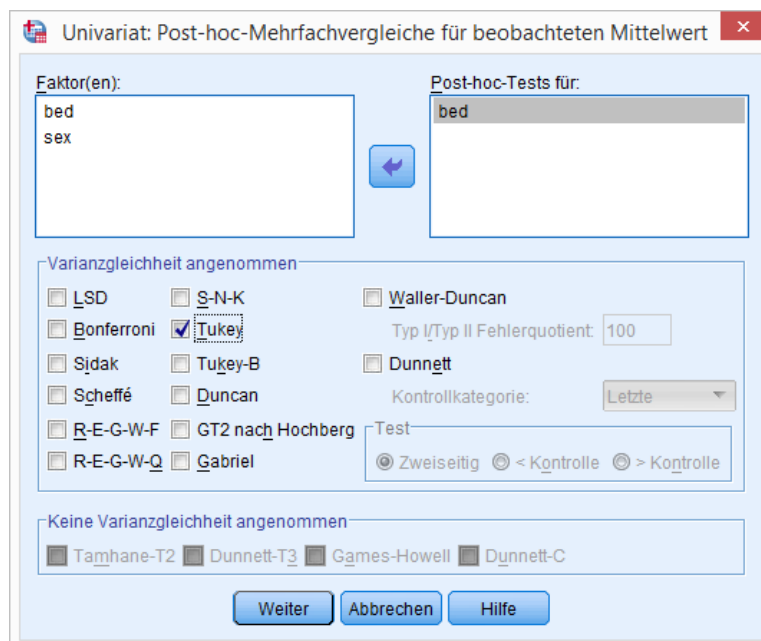
Betätigen Sie das Feld „Optionen“ und aktivieren Sie die Felder „Deskriptive Statistiken“, „Schätzungen der Effektgröße“ und „Beobachtete Trennschärfe“ (siehe SPSS Erläuterungen für Kapitel 5).

## SPSS-Ergänzungen

Rasch, Frieze, Hofmann & Naumann (2014). *Quantitative Methoden. Band 2* (4. Auflage). Heidelberg: Springer.



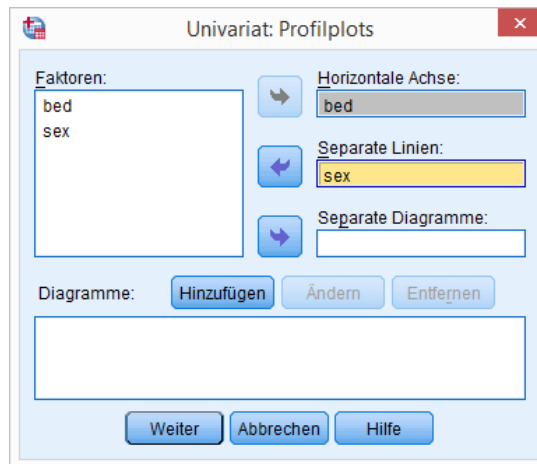
Kehren Sie mit „Weiter“ zu dem vorangegangenen Fenster zurück. Klicken Sie im nächsten Schritt auf „Post Hoc...“. Hier haben Sie die Möglichkeit, aus verschiedenen Verfahren einen geeigneten Post Hoc Test auszuwählen (z.B. Tukey). Ein Post Hoc Test ist erst dann sinnvoll, wenn ein Faktor mehr als zwei Stufen hat, in unserem Beispiel also für den Faktor „Verarbeitungsbedingung“. Ein signifikantes Ergebnis für den Haupteffekt „Geschlecht“ ist dagegen direkt interpretierbar, da hier nur zwei Gruppen miteinander verglichen werden. Leider bietet SPSS keine Möglichkeit, eine Post Hoc Analyse der Wechselwirkung vorzunehmen. Hierzu verwenden Sie bitte das in Kapitel 6.5 beschriebene Vorgehen.



## SPSS-Ergänzungen

Rasch, Friese, Hofmann & Naumann (2014). *Quantitative Methoden. Band 2* (4. Auflage). Heidelberg: Springer.

Obwohl Sie bereits das Anzeigen der Mittelwerte über die Option „Deskriptive Statistik“ aktiviert haben, kann es zusätzlich sinnvoll sein, sich die potentielle Wechselwirkung zwischen den untersuchten Faktoren auch grafisch anzeigen zu lassen. Betätigen Sie hierzu das Feld „Diagramme“. Bewegen Sie die Variable „bed“ in das Feld für „Horizontale Achse“ und die Variable „sex“ in das Feld „Separate Linien“. Klicken Sie auf „Hinzufügen“.



Kehren Sie mit „Weiter“ zu dem vorangegangenen Fenster zurück und starten Sie die Analyse mit OK. Im Folgenden werden die Tabellen des SPSS-Outputs einzeln beschrieben.

Zwischensubjektfaktoren

		Wertbeschriftung	H
Verarbeitungsbedingung	1	strukturell	50
	2	bildhaft	50
	3	emotional	50
Geschlecht	1	maennlich	52
	2	weiblich	98

In der ersten Tabelle des Outputs stehen die untersuchten Faktoren mit ihren einzelnen Stufen bzw. experimentellen Bedingungen. Der Faktor „Verarbeitungsbedingung“ hat drei Stufen („strukturell“, „bildhaft“ und „emotional“), der Faktor „Geschlecht“ zwei Stufen („maennlich“, „weiblich“). In der letzten Spalte sind die Anzahl der Versuchspersonen in den jeweiligen experimentellen Bedingungen aufgelistet.

# SPSS-Ergänzungen

Rasch, Friese, Hofmann & Naumann (2014). *Quantitative Methoden. Band 2* (4. Auflage). Heidelberg: Springer.

## Deskriptive Statistiken

Abhängige Variable: Gesamtzahl erinnerter Adjektive

Verarbeitungsbedingung	Geschlecht	Mittelwert	Standardabweichung	H
strukturell	maennlich	6,38	2,895	16
	weiblich	7,59	3,248	34
	Gesamtsumme	7,20	3,162	50
bildhaft	maennlich	10,07	3,615	15
	weiblich	11,40	4,333	35
	Gesamtsumme	11,00	4,140	50
emotional	maennlich	10,71	4,014	21
	weiblich	12,97	4,153	29
	Gesamtsumme	12,02	4,206	50
Gesamtsumme	maennlich	9,19	4,005	52
	weiblich	10,54	4,498	98
	Gesamtsumme	10,07	4,368	150

Die zweite Tabelle beinhaltet die deskriptiven Statistiken der Faktoren. Die letzte Zeile der Tabelle gibt den Gesamtmittelwert der Anzahl erinnerter Adjektive an, gemittelt über alle Gruppen (10,07 Adjektive). Die beiden vorherigen Zeilen geben die Mittelwerte des Haupteffekts „Geschlecht“ an: Gemittelt über alle Verarbeitungsbedingungen haben Männer 9,19 und Frauen 10,54 Adjektive erinnert. Die Zeilen „Gesamt“ für die einzelnen Stufen des Faktors „Verarbeitungsbedingung“ liefern die bereits aus Kapitel 5 bekannten Werte ohne Berücksichtigung des Faktors Geschlecht: 7,20 erinnerte Adjektive in der strukturellen, 11,00 in der bildhaften und 12,02 in der emotionalen Verarbeitungsbedingung. Die Mittelwerte für die Kombinationen der verschiedenen Faktorstufen lassen sich ebenfalls aus der Tabelle ablesen: So haben z.B. Männer, die die Adjektive strukturell verarbeitet haben, 6,38 Adjektive erinnert, Frauen dagegen 7,59 Adjektive. Die letzte Spalte gibt die Anzahl der Versuchspersonen in den einzelnen Bedingungskombinationen an.

Mit der zweifaktoriellen Varianzanalyse lassen sich drei verschiedene Effektarten untersuchen:

- Gibt es einen systematischen Unterschied zwischen der strukturellen, bildhaften und emotionalen Verarbeitungsbedingung (Haupteffekt „Verarbeitungsbedingung“)?
- Liegt ein systematischer Unterschied zwischen der Erinnerungsleistung von Frauen und Männern vor (Haupteffekt „Geschlecht“)?
- Ist der Einfluss der Art der Verarbeitung der Adjektive anders für Frauen als für Männer (Wechselwirkung „Verarbeitungsbedingung“ x „Geschlecht“)?

Antworten auf diese Fragen liefert die folgende Tabelle:

## Tests der Zwischensubjekteffekte

Abhängige Variable: Gesamtzahl erinnerter Adjektive

Quelle	Typ III Quadratsumme	df	Quadratischer Mittelwert	F	Sig.	Partielles Eta hoch zwei	Dezent. Parameter	Beobachtete Trennschärfe <sup>b</sup>
Korrigiertes Modell	741,623 <sup>a</sup>	5	148,325	10,168	,000	,261	50,840	1,000
Konstanter Term	12976,548	1	12976,548	889,579	,000	,861	889,579	1,000
bed	583,122	2	291,561	19,987	,000	,217	39,975	1,000
sex	85,492	1	85,492	5,861	,017	,039	5,861	,672
bed * sex	7,514	2	3,757	,258	,773	,004	,515	,090
Fehler	2100,570	144	14,587					
Gesamtsumme	18063,000	150						
Korrigierter Gesamtwert	2842,193	149						

a. R-Quadrat = ,261 (Angepasstes R-Quadrat = ,235)

b. Berechnet mit alpha = ,05

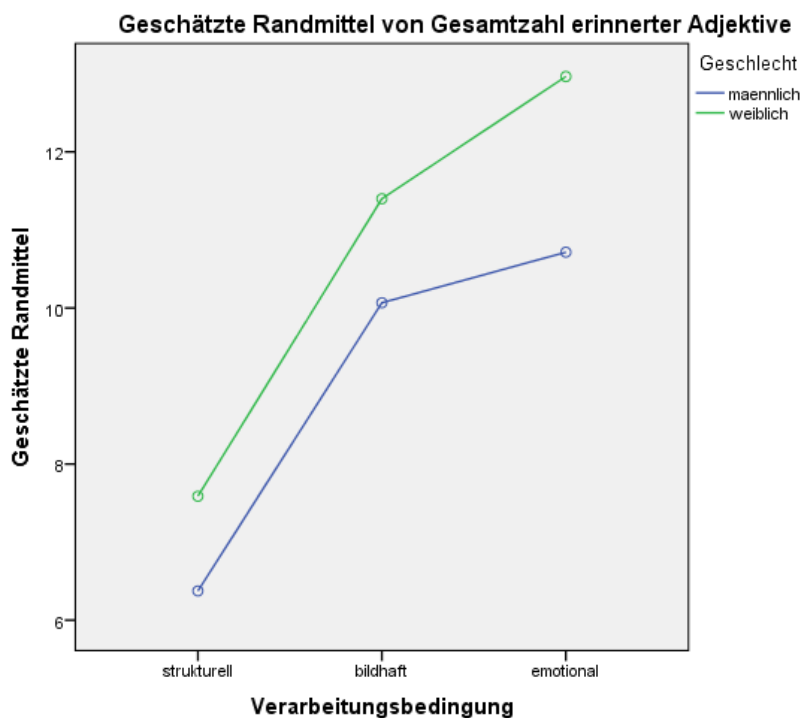
Quelle: <http://www.lehrbuch-psychologie.de/qm>

© Rasch, Friese, Hofmann & Naumann

Die dritte Zeile („bed“) enthält das Ergebnis für den Haupteffekt „Verarbeitungsbedingung“. Der  $F$ -Wert  $F_{(2;144)} = 19,987$  tritt mit einer Wahrscheinlichkeit von  $p < 0,001$  unter der Nullhypothese auf. Wie bereits aus Kapitel 5 bekannt, unterscheidet sich die Erinnerungsleistung stark zwischen den drei experimentellen Bedingungen „strukturell“, „bildhaft“ und „emotional“. SPSS gibt als Effektstärke das partielle Eta-Quadrat an. Es liegt ein großer Effekt vor ( $\eta_p^2 = 0,217$ ; Cohen, 1988). Die Wahrscheinlichkeit, diesen Effekt mit 150 Versuchspersonen zu finden, falls er wirklich existiert, war größer als 99% (Beobachtete Trennschärfe = Teststärke für den empirischen Effekt = 1,000).

Die vierte Zeile zeigt die Werte für den Haupteffekt „Geschlecht“. Die Wahrscheinlichkeit des  $F$ -Werts  $F_{(1;144)} = 5,861$  unter der Nullhypothese beträgt  $p = 0,017$ . Sie ist kleiner als das Signifikanzniveau  $\alpha = 0,05$ . Es besteht also ein signifikanter Unterschied zwischen der Erinnerungsleistung von Frauen und Männern, unabhängig von der Art der Verarbeitungsbedingung. Aus der vorherigen Tabelle der deskriptiven Statistik lässt sich erkennen, dass Frauen deskriptiv mehr Adjektive erinnern als Männer. Aus dem signifikanten Haupteffekt lernen wir nun, dass dieser Unterschied statistisch signifikant ist. Die Effektgröße beschreibt mit  $\eta_p^2 = 0,039$  einen kleinen bis mittleren Effekt. Die Wahrscheinlichkeit, ein signifikantes Ergebnis zu erhalten, falls ein Effekt dieser Größe in der Population existiert, betrug 67,2%.

Die fünfte Zeile gibt die Testung der Wechselwirkung an. Der  $F$ -Wert von  $F_{(2;144)} = 0,258$  ( $p = 0,773$ ) zeigt an, dass die Wechselwirkung nicht signifikant ist. Ist es auf Grund dieses Ergebnisses möglich, die Nullhypothese anzunehmen, dass keine Wechselwirkung zwischen den Faktoren „Verarbeitungsbedingung“ und „Geschlecht“ auf die Erinnerungsleistung besteht? Ein Blick auf das von SPSS erstellte Profildiagramm ist bei der Einschätzung dieser Frage hilfreich.



Die Parallelität der Geraden für Männer und Frauen zwischen den Bedingungen „strukturell“ und „bildhaft“ scheint zu bestätigen, dass zumindest hier keine Wechselwirkung zwischen den beiden Faktoren „Verarbeitungsbedingung“ und „Geschlecht“ besteht. Allerdings erinnern Frauen in der Bedingung „emotional“ mehr Wörter als in der Bedingung „bildhaft“, während dieser Unterschied für Männer geringer ausfällt. Liegt hier wirklich keine Wechselwirkung vor? Für diese Entscheidung ist eine a posteriori Powerberechnung erforderlich (siehe G\*Power Ergänzungen für dieses Kapitel).

Der von SPSS unter „Beobachtete Trennschärfe“ angegebene Wert ist für die a posteriori Powerberechnung in der Regel nicht geeignet. Die „beobachtete Trennschärfe“ gibt die Teststärke an, den in den Daten vorhandenen empirischen Effekt von  $\eta_p^2 = 0,004$  mit der gegebenen Versuchspersonenzahl zu finden. Dieser Effekt ist extrem klein. Und weil kleine Effekte nur mit sehr großen Stichproben mit hoher Sicherheit detektiert werden können, ist die angegebene Teststärke für dieses Effekt ebenfalls sehr klein ( $1-\beta = 0,09$ ).

Heißt das nun, dass wir die Nullhypothese der Wechselwirkung nicht annehmen dürfen? Nein, das heißt es nicht! Denn für die a posteriori Powerberechnung ist es erforderlich, eine inhaltlich relevante Effektgröße festzulegen und zu berechnen, mit welcher Sicherheit wir diese inhaltlich relevante Effektgröße ausschließen können. Sie könnten z.B. festlegen, dass ein mittlerer Effekt nach Cohen (1988) von  $f = 0,25$  ( $\Omega^2 = 0,06$ ) inhaltlich relevant ist. Ist es möglich, in der vorliegenden Untersuchung einen Wechselwirkungseffekt dieser Größe mit ausreichender Sicherheit auszuschließen? Die Teststärke für  $f = 0,25$  liegt bei den vorliegenden Bedingungen für die Wechselwirkung bei  $1-\beta = 78\%$ . Sie können also mit 78%iger Sicherheit ausschließen, dass kein Wechselwirkungseffekt der Größe  $f = 0,25$  vorliegt. Und Sie können sogar mit einer über 99%igen Sicherheit ausschließen, dass ein großer Effekt von  $f = 0,40$  existiert (vgl. G\*Power Ergänzungen zu diesem Kapitel).

Noch einmal: Der von SPSS unter „Beobachtete Trennschärfe“ berichtete Wert ist nur in wenigen Fällen aussagekräftig. Warum? Bei einem signifikanten Ergebnis ist die Berechnung der Teststärke nicht notwendig, da ja ein Effekt gefunden wurde. Es ist also uninteressant, wie groß die Wahrscheinlichkeit war, einen Effekt dieser Größe zu finden, falls er existiert, denn auf Grund des signifikanten Ergebnisses hat man bereits die Entscheidung getroffen, dass er existiert.

Bei nicht signifikanten Ergebnissen ist die „Beobachtete Trennschärfe“ ebenfalls wenig hilfreich, weil der Wert sinkt, je kleiner der empirische Effekt bzw. je kleiner der Unterschied zwischen den experimentellen Bedingungen ist. Besonders in solchen Fällen, in denen die Wahrscheinlichkeit des  $F$ -Werts unter der Nullhypothese sehr hoch ist, und wir eigentlich besonders sicher sein sollten, dass die Nullhypothese zutrifft, gibt die „Beobachtete Trennschärfe“ einen sehr niedrigen Wert für die Teststärke an. Deshalb erlaubt die „Beobachtete Trennschärfe“ keine inhaltlich relevante Aussage über die Teststärke einer Untersuchung. Stattdessen muss für die a posteriori Teststärkeberechnung ein inhaltlich relevanter Effekt festgelegt werden. Allein diese Berechnung erlaubt eine sinnvolle Aussage über die Sicherheit, nach einem nicht signifikanten Ergebnis einen inhaltlich relevanten Effekt auszuschließen.

Eine Ausnahme für eine sinnvolle Nutzung der „Beobachteten Trennschärfe“ tritt auf, wenn der aus den vorliegenden Daten geschätzte empirische Effekt eine inhaltlich relevante Größe erreicht, und trotzdem kein signifikantes Ergebnis auftritt. In diesem Fall macht die „Beobachtete Trennschärfe“ eine inhaltlich sinnvolle Aussage über die Wahrscheinlichkeit, mit der ein Effekt dieser Größe in der vorliegenden Untersuchung gefunden werden konnte. In diesen Fällen wird die beobachtete Teststärke eine unbefriedigende Größe erreichen.

## SPSS-Ergänzungen

Rasch, Friese, Hofmann & Naumann (2014). *Quantitative Methoden. Band 2* (4. Auflage). Heidelberg: Springer.

Die Tabelle mit der Post Hoc Analyse entspricht im wesentlichen der in den SPSS Ergänzungen für Kapitel 5 beschriebenen Ausgabe: Während sich die Mittelwerte zwischen den Bedingungen „strukturell“ und „bildhaft“ signifikant unterscheiden, zeigt sich kein Unterschied zwischen den Bedingungen „bildhaft“ und „emotional“.

### Post-Hoc-Tests

#### Verarbeitungsbedingung

##### Mehrfachvergleiche

Abhängige Variable: Gesamtzahl erinnerter Adjektive

Tukey-HSD

(I) Verarbeitungsbedingung	(J) Verarbeitungsbedingung	Mittelwertdifferenz (I-J)	Standardfehler	Sig.	95 % Konfidenzintervall	
					Untergrenze	Obergrenze
strukturell	bildhaft	-3,80*	,764	,000	-5,61	-1,99
	emotional	-4,82*	,764	,000	-6,63	-3,01
bildhaft	strukturell	3,80*	,764	,000	1,99	5,61
	emotional	-1,02	,764	,378	-2,83	,79
emotional	strukturell	4,82*	,764	,000	3,01	6,63
	bildhaft	1,02	,764	,378	-,79	2,83

Basierend auf beobachteten Mittelwerten.

Der Fehlerterm ist Mittel der Quadrate(Fehler) = 14,587.

\*. die Mittelwertdifferenz ist auf der Stufe ,05 signifikant.

## Homogene Untergruppen

### Gesamtzahl erinnerter Adjektive

Tukey-HSD

Verarbeitungsbedingung	H	Subset	
		1	2
strukturell	50	7,20	
bildhaft	50		11,00
emotional	50		12,02
Sig.		1,000	,378

Mittelwerte für Gruppen in homogenen Subsets werden angezeigt.

Basierend auf beobachteten Mittelwerten.

Der Fehlerterm ist Mittel der Quadrate(Fehler) = 14,587.

a. Verwendet harmonischen Mittelwert des Stichprobenumfangs = 50,000.

b. Alpha = ,05.

## Effektstärke und empirische Teststärke einer zweifaktoriellen Varianzanalyse ohne Messwiederholung

Die Effektstärke  $\eta_p^2$  sowie die empirische Teststärke („Beobachtete Trennschärfe“) werden von SPSS im Output mit angegeben, wenn die entsprechenden Optionen zuvor aktiviert wurden (siehe oben). Darüber hinaus gibt es die Möglichkeit, in G\*Power über die Varianzen die empirische Effektstärke auszurechnen (siehe G\*Power-Ergänzungen zu Kapitel 6 sowie Kapitel 6.3 in Band 2).



### Literatur

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NY: Erlbaum.