

Fragebogen- Evaluierung



ISPA consult GmbH
Endersbacher Str. 57
71334 Waiblingen
T. +49 (0)7151 173311-0
info@ispa-consult.de
www.ispa-consult.de

Fragebogen-Evaluierung

Erfolgreiche Gestaltung eines Fragebogens nach Standardvorgaben

Inhaltsverzeichnis

1. Qualitätssicherung durch Fragebogenevaluation	3
2. Fragebogenformulierung, Fragebogenkonstruktion und Themenfelder	4
3. Messung der Objektivität	6
4. Messung der Reliabilität	7
5. Messung der Validität	9
6. Abgrenzung Validität und Reliabilität	10
7. Kontexteffekte und Kognitive Urteilsbildung	11

1. Qualitätssicherung durch Fragebogenevaluation

Die ISPA consult GmbH evaluiert eingereichte Fragebögen von Firmenkunden durch gründliche Prüfung und Bewertung mit Verbesserungsvorschlägen. Der Fragebogen muss anhand der Zielsetzung der Befragung entwickelt werden und sollte alle Fragen und Items enthalten, die später ausgewertet werden sollen.

Um eine fundierte Aussage machen zu können, ob ein vorliegender Fragebogen **gut** oder **verbesserungsbedürftig** ist, sind folgende Prüfungen erforderlich:



Die praktische Durchführung dieser Prüfungen ist Aufgabe unserer „MAB-Werkstatt“. Sie ist auf Prüfung und Verbesserung von eingereichten Fragebögen der Kunden spezialisiert und hilft Ihnen, die Qualität Ihrer Mitarbeiterbefragung zu verbessern.

GEPRÜFT



2. Fragebogenformulierung, Fragebogenkonstruktion und Themenfelder

2.1 Die „10 Gebote der Fragebogenformulierung“

Für die Fragebogenevaluation werden als erstes die Fragen-Formulierungen anhand der „10 Gebote der Fragebogenformulierung“ nach R. Porst 2000 (Question Wording - Zur Formulierung von Fragebogen-Fragen. ZUMA How-to-Reihe, Nr. 2) analysiert.

Die „10 Gebote der Fragebogenformulierung“ haben zum Inhalt:

- (1) Einfache, unzweideutige Begriffe verwenden, die von allen Befragten in gleicher Weise verstanden werden!
- (2) Unklare Begriffe definieren!
- (3) Fragen vermeiden, die auf Informationen abzielen, über die viele Befragte mutmaßlich nicht verfügen!
- (4) Lange und komplexe Fragen vermeiden!
- (5) Hypothetische Fragen vermeiden!
- (6) Unterstellungen und suggestive Fragen vermeiden!
- (7) Doppelte Stimuli und Verneinungen vermeiden!
- (8) Fragen mit eindeutigem zeitlichem Bezug verwenden!
(z.B. nicht: in letzter Zeit, früher, ...)
- (9) Die theoretische Angemessenheit der Antwortkategorien überdenken!
- (10) Antwortkategorien verwenden, die erschöpfend und disjunkt sind!

2.2 Prüfung der Fragebogenkonstruktion

Der Fragebogenformulierung folgend wird die Fragebogenkonstruktion (Frageformen, Fragetypen, Antwortkategorien) genauer unter die Lupe genommen. (Quelle: Sozialwissenschaftliche Erhebungsmethoden, Vorlesungsreihe der Universität Stuttgart).

a) **Frageformen.** Zu unterscheiden sind

- Geschlossene und offene Fragen (Strukturtypen)
- Kontrollfragen
- Filterfragen, Gabelungen
- Trichterfragen
- Warming-Up-/ „Eisbrecher“- / Kontakt- / Einleitungs-/ Eröffnungs- / Übungs- / Trainingsfragen (Aufbau von Spannungskurven, „Heikle“ Fragen am Schluss etc.)
- Ablenkungs- / Pufferfragen
- Layout Ästhetik etc.

b) **Fragetypen.** Fragen können sich beziehen auf



- Einstellungen, Meinungen, allg. Bewertungen (z. B. Wichtigkeit, ...)
- Werthaltungen
- Überzeugungen (normative und kognitive beliefs)
- Verhalten, Verhaltensabsicht
- Eigenschaften der Befragten (sozioökonomisch und - demographisch)
- Gefühle (affektive Komponente)

c) **Antwortkategorien.** Zu prüfen ist zum Beispiel



- Das Verhältnis von positiv und negativ formulierten Antworten
- Die Qualität der Antworten (präzise, disjunkt, überschneidungsfrei, erschöpfend etc.)
 - Weiß-nicht-Kategorien und was damit erfasst werden soll explizit (mehr missings) vs. implizit (mehr Nonattitudes)
 - Die theoretische Angemessenheit der Antwortkategorien (muss stets bedacht werden)
 - Das Vorhandensein von Skalendrehungen (nicht die Antwortkategorien „drehen“, sondern die Item-Formulierung!)
- Numerische und semantische Kategorienbezeichnungen
- Die Art der verwendeten Skalierung (z.B. gerade oder ungerade Antwortkategorienzahl)

2.3 Prüfung der Themenfelder

- Prüfung auf Aktualität (Lückentest)
- Analyse der Anzahl der Items (Frage - Antwortmöglichkeiten; auch bezogen auf die Anzahl der Fragen pro Themenfeld; Prüfung auf Ausgewogenheit)

3. Messung der Objektivität

Bei der Messung der Objektivität geht es um die Unabhängigkeit der Testergebnisse vom Testanwender. Dabei sind zu unterscheiden:

a) **Durchführungsobjektivität**

Konstanz der Untersuchungsbedingungen - Durch eine möglichst hohe Standardisierung können Störfaktoren wie Item-Reihenfolge, Fragebogen-Unterbrechung etc. vermieden werden.

b) **Auswertungsobjektivität**

Die Reaktion der Befragten auf eine Frage kann oft nur unzureichend als "Kreuzchen" im Fragebogen abgebildet werden. Vor allem die Codierung offener Fragen ist fehleranfällig. Fehler in der Dateneingabe stellen ebenfalls eine Beeinträchtigung der Auswertungsobjektivität dar. Um die Auswertungsobjektivität quantitativ zu bestimmen kann man Befragungen von zwei verschiedenen Auswertern codieren lassen. Die mittlere Korrelation kann dann als Maß der Auswertungsobjektivität interpretiert werden (Hinweis: Die Auswertungsobjektivität wird von der ISPA consult GmbH nicht geprüft, ist jedoch bei Bedarf und unter bestimmten Voraussetzungen bei Wiederholungsbefragungen realisierbar).

c) **Interpretation der Ergebnisse**

Hier ist eine hohe Objektivität gewährleistet, wenn verschiedene Interpretatoren mit vergleichbarem Wissensstand zum selben Ergebnis bei der Beurteilung der Befragungsergebnisse kommen. Interpretationshinweise und Untersuchungsdetails müssen dafür zur Verfügung gestellt werden, Normwerte oder Benchmarks ermöglichen eine quantitative Interpretation der Ergebnisse.

Die Objektivität kann und sollte das Unternehmen vor der Wiederholungsbefragung sicherstellen.

4. Messung der Reliabilität

Die Reliabilität bezeichnet die Messgenauigkeit bzw. Zuverlässigkeit eines Verfahrens, also beispielsweise die Genauigkeit, mit der eine Skala ein Merkmal misst. Eine Reliabilität von 1 ist erreicht, wenn die Messergebnisse exakt repliziert werden können. Die Reliabilität wird über die Berechnung von Korrelationen zwischen Messwerten ermittelt. Bei angenommener großer Merkmalsstabilität können Reliabilitätskoeffizienten in der Größenordnung um .80 bereits als gut gelten, ein Wert von 0.70 ist befriedigend (z.B. Test von Kendalls Tau).

Eine generelle Antwort kann hier nicht gegeben werden, da viele Einflussfaktoren zu berücksichtigen sind. So sind z.B. Untersuchungszeitpunkt (Retest-Design), inhaltliche Vergleichbarkeit der Items sowie die Anzahl der Items und Skalenwerte zu nennen.

Je nachdem, welche Messwerte miteinander korreliert werden, unterscheidet man verschiedene Arten von Reliabilitäten:

a) Paralleltest-Methode

Die Ergebnisse zweier parallel konstruierter Fragebögen, die man denselben Personen vorgegeben hat, werden miteinander korreliert.

Korrelation 1	Korrelation 2
Person A	Person A
Zeitpunkt 1	Zeitpunkt 1
Fragebogen 1	Fragebogen 2

b) Retest-Reliabilität

Die Korrelation zwischen zwei wiederholten Messungen bei den gleichen Personen zu unterschiedlichen Zeitpunkten.

Korrelation 1	Korrelation 2
Person A	Person A
Zeitpunkt 1	Zeitpunkt 2
Fragebogen 1	Fragebogen 1

Durch den zeitlichen Abstand der 2 Untersuchungen sollen Erinnerungseffekte bei der befragten Person gemindert werden. Dieses Verfahren wird von der ISPA consult GmbH angewendet, wenn man vermutet, dass sich die Einstellung der TeilnehmerInnen der MAB im Untersuchungszeitraum vermutlich nicht ändert. Wertvorstellungen und generelle Einstellungen sind hier abfragbar, die Frage nach Stimmungen dagegen ist für diesen Test ungeeignet.

c) Split-half-Reliabilität

Dabei teilt man die Items eines Fragebogens in zwei Gruppen und korreliert die Ergebnisse beider Testhälften. Die beiden Teile beleuchten dasselbe Merkmal mit unterschiedlichen Items. Würde man im zweiten Teil ein anderes Merkmal einschätzen als im ersten, würde es zu geringen Korrelationen kommen.

Korrelation 1	Korrelation 2
Person A	Person A
Zeitpunkt 1	Zeitpunkt 2
Fragebogen 1, Teil 1	Fragebogen 1, Teil 2

Bei der Aufteilung der Items auf die beiden Teile können verschiedene Strategien gewählt werden. Neben einer einfachen Halbierung der Items und Verteilung auf die 2 Teile kann auch eine Aufteilung nach Itemkennwerten erfolgen. Dazu werden Itempaare anhand von Mittelwert, Streuung und Korrelation mit dem Gesamtindex gebildet und dann auf die beiden Fragebogenteile verteilt. Beim Split-half-Verfahren fällt die Reliabilität im Vergleich zum Paralleltestverfahren geringer aus, da weniger Items korreliert werden. Dies kann mit der Spearman-Brown-Formel für Testverdoppelung korrigiert werden.

5. Messung der Validität

Unter der Validität versteht man das Ausmaß, in dem der Test das misst, was er messen soll.

a) Logische Validität

Logische Validität meint so etwas wie eine triviale Gültigkeit eines Erhebungsverfahrens. Sie wird auch inhaltliche Validität oder Augenschein-Validität genannt und ist nicht sehr brauchbar.

b) Externe Validität

Externe Validität setzt voraus, dass eine andere Messung dessen, was der Fragebogen messen soll, verfügbar ist. Die Validität wird über die Korrelation der Testergebnisse mit dieser anderen Messung desselben Merkmals berechnet. Ein Beispiel ist hier der Vergleich von Wahlprognose und Wahlergebnis.

c) Interne Validität

Die Ergebnisse einer Testkonstruktion müssen letztlich mit dem Gesamtwissen zum Konstrukt übereinstimmen. Ein Verfahren zur Bestimmung der internen Validität ist die Faktorenanalyse.

d) Konstruktvalidität

Bezeichnet in der multivariaten Statistik einen Teilaspekt der Validität und liegt vor, wenn die Messung eines Konstrukts weder durch systematische Fehler noch durch andere Konstrukte verfälscht ist. Es geht gewissermaßen um die Fragestellung, ob sich die zur Messung des Konstrukts gewählten Indikatoren auf eine Weise verhalten, als dass man sie gemeinsam als ein „intellektuelles Ganzes“ auffassen kann.

6. Abgrenzung Validität und Reliabilität

Unter der Validität versteht man das Ausmaß, in dem der Test das misst, was er messen soll.

Die Abgrenzung der Validität gegenüber der Reliabilität verdeutlicht das nachfolgende Beispiel:

Angenommen: Gegenstand der Befragung sei die Messung der Mitarbeiterzufriedenheit. Eine niedrige Mitarbeiterzufriedenheit gilt als unerwünscht und wird daher ungern preisgegeben, insbesondere dann, wenn diese Information mittels eines Analyse-instrumentes (Fragebogen) eingeholt wird.

Ein Teil der Befragten könnte dazu neigen, in einer Befragung stets die Beantwortung der Zufriedenheit auf ein bestimmtes Maß zu beschränken, egal wie oft die Befragung durchgeführt wird. Dadurch kann das Instrument zwar eine hohe Reliabilität besitzen (weil stets die gleiche Mitarbeiterzufriedenheit gemessen wird), aber es besitzt keine Validität, denn die gemessene Zufriedenheit weicht unter Umständen von der wahren Mitarbeiterzufriedenheit und dem tatsächlichen Betriebsklima ab. Es ist folglich möglich, dass Messwiederholungen stets dasselbe Ergebnis erbringen, aber die Messungen etwas anderes messen, als beabsichtigt ist, und daher keine Validität besitzen. Umgekehrt ist es nicht möglich, dass ein Instrument valide Daten liefert ohne gleichzeitig zuverlässig zu sein. Ein Instrument ist umso valider, je weniger systematische Fehler vorkommen.



7. Kontexteffekte und Kognitive Urteilsbildung

(allg. auch „Reponse-Effekte“ genannt)

Antwortverzerrungen z.B. „der Halo-Effekt“ kann auch im Fragebogen auftreten. Damit ist gemeint, dass einzelne Fragen andere „überstrahlen“ können. Wenn beispielsweise die vorhergehende Frage bestimmte Gedanken oder Gefühle auslöst, kann dies Auswirkungen auf die Antwort der nächsten Frage haben. Der Effekt entsteht auch bei mangelnder Motivation und Informiertheit der Beurteilenden. Der Halo-Effekt muss daher ebenso bei der Fragebogenevaluation beachtet und geprüft werden.

Der **Fragekontexteffekt** (auch: Reihenfolge-Effekt, Fragereiheneffekt oder Positionseffekt) besagt, dass die Reihenfolge der gestellten Frage in einem Fragebogen einen Einfluss auf die Interpretation und die Bewertung der Fragen seitens des Befragten haben kann. Die Beantwortung einer Frage kann wesentlich durch die vorangegangenen Fragen beeinflusst sein. Es handelt sich somit um eine systematische Verzerrung. Um einen solchen Fragekontexteffekt zu identifizieren, kann die Reihenfolge der Fragen im Pretest oder in einem Split ballot (gegabelte Befragung) variiert werden. Dadurch erhöht sich jedoch auch der Aufwand in der Fragebogenkonstruktion (wird deshalb nicht von der ISPA consult GmbH durchgeführt, da der Fragebogen schon im Einsatz ist).

Mit **Assimilationseffekten** ist zu rechnen, wenn spezielle vor allgemeine Fragen zum selben Themenkomplex gestellt werden. Auch dieser Aspekt wird geprüft.