

EXPERTENBEITRAG 01 | 2015

Ingo Radermacher
Der Entscheidungsphilosoph®

INGO RADERMACHER | ingo@entscheidungsphilosoph.de

Happy sheets – oder: „Wie war ich, Schatz?“

Der Feedbackbogen zwischen Sinnlosigkeit
und zarten Ansätzen zur Verbesserung

Inhalt

Feedbackbogen: Wirklich ein must-have?	4
Ziel: Wenn überhaupt, dann besser fragen.	4
„Happy sheets...“	5
Besser konstruiert	6
KONSTRUKTIONSTIPP 1:	
Warum fragst Du?	6
KONSTRUKTIONSTIPP 2:	
Bist Du aufrichtig?	7
KONSTRUKTIONSTIPP 3:	
Wie soll ich fragen?	8
KONSTRUKTIONSTIPP 4:	
Was ist der geeignete Fragetypus?	9
KONSTRUKTIONSTIPP 5:	
Welche Skala ist geeignet?	10
KONSTRUKTIONSTIPP 6:	
Passt die Skala zur Fragestellung?	13
KONSTRUKTIONSTIPP 7:	
Orientierung am Befragten!	14
Auswertung von Fragebögen:	
Unsinn mit statistischer Genauigkeit	15
AUSWERTUNGSTIPP 1:	
Hast Du die Toten berücksichtigt?	15
AUSWERTUNGSTIPP 2:	
Sind die Antworten vielleicht sozial erwünscht?	16
AUSWERTUNGSTIPP 3:	
Median statt/und Mittelwert	16
AUSWERTUNGSTIPP 4:	
Erfolgszuschreibungen mit maximaler Genauigkeit	17
AUSWERTUNGSTIPP 5:	
War nicht völlig egal, was gefragt wurde?	18
Fazit	19
Anmerkungen	19
Literaturverzeichnis	19
Über den Autor	20



INGO RADERMACHER | ingo@entscheidungsphilosoph.de

Happy sheets – oder: „Wie war ich, Schatz?“

**Der Feedbackbogen zwischen Sinnlosigkeit
und zarten Ansätzen zur Verbesserung**

Zur Zeit des Christoph Kolumbus glaubten die Menschen, die Welt sei eine Scheibe. Kolumbus wurde daher prophezeit: Er werde bei seiner Reise ans Ende der Welt vom Rand dieser Scheibe abstürzen. Dies ist unsere gängige Annahme bezüglich des Weltbildes der Menschen im Mittelalter – und diese Annahme ist ein Irrtum!

Denn Untersuchungen seit den 1990er Jahren zeigen:¹ Schon in der Antike, zur Zeit des Platon und Aristoteles, gab es die Vorstellung, die Welt sei nicht etwa eine Scheibe, sondern eine Kugel. Im europäischen Mittelalter und früher Neuzeit – die Zeit des Christoph Columbus – wurde die Kugelgestalt der Erde kaum noch in Frage gestellt. Die moderne Fehlannahme, der mittelalterliche Mensch habe an eine scheibenförmige Erde geglaubt, fand erst im 19. Jahrhundert Verbreitung – vor allem durch einen Spaß des Satirikers Washington Irving in seiner Erzählung „Das Leben und die Reisen des Christoph Kolumbus“.²

Dennoch: Behauptet man heute, im Mittelalter habe die Vorstellung einer scheibenförmigen Erde vorgeherrscht, wird man selten bis nie korrigiert. Stattdessen erntet man zustimmendes Kopfnicken.

Feedbackbogen: Wirklich ein must-have?

Ähnlich verhält es sich mit Fragebögen zur Evaluation von Seminaren, Trainings, Vorträgen oder anderen Weiterbildungsveranstaltungen. Ihr Sinn – oder Unsinn – wird nicht hinterfragt. Sie sind per se gut, richtig und wichtig. Denn: Das macht man heute so! Das macht jeder so! – „Feedbackbögen?“ Einhelliges zustimmendes Kopfnicken...

Entsprechend herrscht eine Feedbackbogen-Epidemie. Man kann sich ihrer gerade in der Weiterbildung mittlerweile kaum mehr erwehren. Feedbackbögen werden inflationär verteilt und ihre Beantwortung eingefordert – mindestens im Nachgang per E-Mail (selbstverständlich mit penetranter Mehrfach-Erinnerung bei Nicht-Antwort).

Aber: Ist dieses Vorgehen nun richtig? Wie steht es um Sinn und Unsinn der Feedbackbögen? Nutzen sie jemandem?

Denn: Fragen kostet! Jede Erstellung, Beantwortung, Nachfassaktion und vor allem auch die Auswertung eines Fragebogens hat ihren Preis. Sie kostet Lebenszeit, Energie, Aufmerksamkeit – allesamt rare Güter.

Im Folgenden werde ich die Praxis der Feedbackbögen in der Weiterbildung kurz unter der Fragestellung reflektieren: Sind Fragebögen sinnvoll? Vor allem werde ich dann Tipps zur Konstruktion von Fragebögen und daran anschließend zu deren Auswertung geben. Handlungsleitend ist dabei die Devise: Verbesserung der Fragebogengüte durch einfache Maßnahmen.

Ziel: Wenn überhaupt, dann besser fragen.

Die Ausführungen stützen sich auf Beispiele aus der Feedbackbogen-Praxis;³ an ihnen werden die thematisierten Schwierigkeiten und Fehler illustriert. Weiter liegt mir daran, an Modellfragen exemplarisch aufzuzeigen, wie Verbesserungen konkret aussehen können.

Ich verwende die Begriffe „Feedbackbogen“, „Fragebogen“ und „Evaluationsbogen“ synonym – weitere begriffliche Abgrenzungen und Einordnungen sind in diesem Kontext entbehrlich. Es geht um den – auf den ersten Blick standardisierten – Fragebogen, mittels dessen Teilnehmer im Anschluss an eine Bildungsmaßnahme nach ihrer Einschätzung des Seminars und seiner Randbedingungen (z.B. Veranstaltungsort) befragt werden. In methodischer Hinsicht handelt es sich dabei im Grundsatz zumeist um einen (oft sicher unbewussten) Methodenmix einer im Kern aber quantitativen Evaluation eines speziellen Weiterbildungsereignisses.

Es geht mir nicht um die wissenschaftliche Seite; hierzu sei verwiesen auf den Markt an Literatur zur Fragebogenkonstruktion und -auswertung und das große Spektrum an Fachartikeln zu Einzelaspekten, aus dem sich bei Bedarf bedient werden kann.



Wobei zu berücksichtigen ist, dass die überwiegende Zahl dieser Publikationen eine Evaluation im ausschließlich wissenschaftlichen Kontext – und eben nicht im Weiterbildungskontext – zum Thema hat. Demgegenüber zielt mein Blick auf ausgewählte konkrete Praxisbeispiele aus dem Bereich der Weiterbildung und nachvollziehbare Anregungen zur Verbesserung: um praxisrelevante Tipps sowohl zur Fragebogengenerierung als auch zu Auswertung.

„Happy sheets...“

Wer schon mal einen Feedbackbogen zu einem Seminar oder Vortrag ausfüllen musste, weiß um den Eindruck, dass es sich um einen eher unstrukturierten, zuweilen nicht wirklich ernst gemeinten, zufällig entstandenen und/oder abgeschrieben Fragebogen handelt. Gebeten wird nicht nur um Feedback zum Seminar selbst, zum Vortrag und Referenten, sondern auch zu Rahmenbedingungen wie die Rechtzeitigkeit der Seminareinladung und auch zu Aspekten wie Catering oder Raumtemperatur. Daher hat diese Art und Weise der Befragung den einschlägigen Fragebögen auch den Terminus „*Happy sheets*“ eingebracht – oder auch „*Smile-Sheets*“ oder „*Happiness-Sheets*“; „[...] im Allgemeinen gehen die Frager davon aus, dass das Ende eines Seminars ein günstiger Zeitpunkt ist, besonders ‚glückliche‘ Antworten zu erhalten.“⁴ Und das, wo jeder professionelle Trainer, Referent oder Speaker eh weiß: Wenn ich will, dann bekomme ich nahezu jede Teilnehmergruppe dazu, mir das Feedback zu geben, dass ich möchte – insbesondere am Veranstaltungsende.

Besser konstruiert

Erklärtes Anliegen der Fragebögen ist im Normalfall: (zumeist unmittelbar) die Reaktion von Teilnehmenden auf ein Seminar oder eine Weiterbildungsveranstaltung festzuhalten, etwa im Sinne einer Zufriedenheits- und/oder Lernerfolgskontrolle. Tatsächlich sind „Happy sheets“ oftmals nicht mehr als eine Alibi-Übung, die zumindest nach einer Erfolgskontrolle oder -messung aussieht. Denn für eine wirksame Erfolgskontrolle bräuchte es ein methodisches Fundament, wenigstens die Konzeption einer neutralen Einheit der Lernerfolgsmessung und die Festlegung von Bewertungskriterien. Nicht selten gibt es auf Seiten der Weiterbildungsanbieter auch ein Qualitätsmanagement; ein Seminar-Feedback-Fragebogen kann dann ein Bestandteil des Qualitätsmanagements sein, etwa des Bildungscontrollings. Gerade hierfür braucht es Zahlen, Daten, Fakten – messbar, vergleichbar. Außerdem gibt es, wenn es ausschließlich um die Lernerfolgskontrolle geht, andere Mittel und Wege; diese sind zwar aufwendiger, sind aber geeignet(er), wirklich robuste Messergebnisse und verlässliche(re) Aussagen über Lernerfolg und Kompetenzerwerb zu treffen, als „Happy sheets“.⁵

Tatsächlich beschleicht einen angesichts dieser Fragebögen oftmals der Eindruck: Zufriedenheits- oder Lernerfolgskontrolle ist gar nicht das Ziel; stattdessen hat der „Happy sheet“ ganz andere Beweggründe.

KONSTRUKTIONSTIPP 1:

Warum fragst Du?

„Ask a question – and you ’ll get an answer. But it won ’t tell you much“.⁶ Beispiel gefällig? „Und, wie war ich, Schatz?“ – oder aus einem Seminar-Fragebogen:

Wie hat Ihnen unser Referent gefallen?

Es gibt intelligente Fragen, und es gibt weniger intelligente Fragen, und viele „Happy sheets“ geben Beispiele für Fragen, auf die es keinerlei intelligente Antworten gibt.

Das Problem liegt meist bereits im Beginn einer Fragebogenkonstruktion: Es fehlt ein klares Ziel – am besten formuliert in einer übergreifenden Frage. Denn nur dann ist es möglich, sowohl die Relevanz einzelner Fragen als auch die Stringenz der Fragen mit Blick auf die Zielerreichung zu prüfen. Mangelt es an einer solchen klaren Zielkonzeption, erhält man statt Antworten wortreiches Geschwätz.



FAZIT: Formuliere ein klares Ziel für die Befragung und prüfe jede Frage auf ihren Beitrag zur Zielerreichung.

KONSTRUKTIONSTIPP 2:

Bist Du aufrichtig?

Weiter gilt: Die Befragten sind nicht dumm, beschränkt, naiv oder gänzlich unkritisch. Geht es bei einem Fragebogen nur vordergründig um Zufriedenheits- oder Lernkontrolle, tatsächlich aber um Marketingmaßnahmen, um die Ermittlung von Akquisepotenzial, die Adresssammlung oder die Gewinnung von Referenzen für die Website und Social Media, so wird dieses „Segeln unter falscher Flagge“ auch als solches wahrgenommen werden! Und genau das ist das Problem bei nicht wenigen Fragebögen. Beispiele?

Ich habe Interesse an Beratung: ja / nein

zum Thema: _____

Anzahl der Mitarbeiter: _____

Anschriften von Firmen und Personen, die in Zukunft eingeladen werden sollten.

Dürfen Auszüge aus Ihrem Feedback als Teilnehmerstimme auf meiner Homepage veröffentlicht werden?

- Ja, mit Foto und Link zur Homepage
- Ja, mit Name und Firma
- Ja, mit Name
- Ja, ohne Name
- Nein

Grundsätzlich können natürlich auch solche Fragen gestellt werden – doch dann nicht unter der Überschrift „Ihre Meinung ist uns wichtig“. Denn sonst bewahrheitet sich leicht das geflügelte Wort Goethes: „Man merkt die Absicht, und ist verstimmt“.



FAZIT: Sei über den Grund der Befragung ehrlich zur Befragungsperson.

Lege ein Ziel für den Fragebogen fest. Formuliere dieses Ziel (Beispiel: „Der Seminar-Feedback-Fragebogen hat zum Ziel, Kenntnisse über die Einschätzung der Teilnehmer zur Seminarqualität zu erlangen.“). Prüfe jede Frage darauf, inwieweit sie tatsächlich relevant für dieses Befragungsziel ist.

KONSTRUKTIONSTIPP 3:

Wie soll ich fragen?

Zur Beantwortung eines Fragebogens wird dem Befragten relativ viel abverlangt:

1. die gestellte Frage verstehen,
2. die zur Beantwortung relevanten Informationen aus dem Gedächtnis abrufen,
3. ein Urteil bilden und dieses in das Antwortformat einpassen.

Kleinste Details der Frageformulierung sind hier extrem wichtig.

So ist auf die Frage – so man sie denn überhaupt so nennen darf:

Ihr Seminarfeedback zu den Methoden:

ein bunter Blumenstrauß an Antworten erwartbar. Gleichzeitig ist unklar, wem die Antworten wozu nützen werden.

Gegenbeispiel:

Im Seminar kamen u.a. folgende Methoden zum Einsatz:

a. Blitzlicht, **b.** Kleingruppenarbeit, **c.** Rollenspiel, **d.** Metaplan, **e.** Stationenarbeit.

A. Welche dieser Methoden fanden Sie für sich sehr hilfreich – und warum?

B. Welche dieser Methoden fanden Sie für sich weniger hilfreich – und warum?

C. Welche Methode haben Sie in diesem Seminar vermisst? Warum hätten Sie sie in diesem Seminar hilfreich gefunden?

Sprachliche und logische Mängel trüben die klare Beantwortung. So hat ein „oder“ in der Frage eines Feedbackbogens grundsätzlich nichts verloren. Denn es macht eine eindeutige und zielsichere Beantwortung einer Frage nahezu unmöglich.

Was hat mich an diesem Seminar geärgert, enttäuscht, frustriert?
Oder auch verblüfft/irritiert/nachdenklich gemacht?

Wenn für eine Fragestellung zwei Begriffe (oder mehr) zur Auswahl stehen, dann wähle den einfachsten und unzweideutigsten. Denn Ziel ist es, dass der Fragebogen von allen in gleicher Weise verstanden wird. Hierzu sollten Fragen kurz und eindeutig sein sowie einen eindeutigen zeitlichen Bezug haben („in der letzten Woche“ statt „in der letzten Zeit“). Ferner sind hypothetische Fragen zu vermeiden:

Stellen Sie sich vor, sie hätten alles, was Sie sich nach diesem Seminar vorgenommen haben, umgesetzt. Welche weiteren Seminarinhalte wären danach für Sie interessant?

Gleiches gilt für die Verwendung von Suggestivfragen – dass daraus kein Erkenntnisgewinn folgt, ist eigentlich klar, nur offensichtlich nicht jedem.

Wie gut stand ich mit Ihren Zuhörern in Kontakt?

Es sollte grundsätzlich nur das abgefragt werden, was für das Ziel der Befragung relevant ist – denn: Vieles mag interessant sein, bleibt aber unwesentlich für die Auswertung.



FAZIT: Frage klar und eindeutig zum Hier und Jetzt und (nur) das, was Du mit Blick auf Dein Befragungsziel wirklich wissen musst.

KONSTRUKTIONSTIPP 4:

Was ist der geeignete Fragetypus?

An klassischen Fragetypen unterscheidet man bekanntlich: die geschlossene Frage mit Einfachnennung, die geschlossene Frage mit Mehrfachnennung, die offene Frage, die halboffene Frage.

Die ersten drei Fragetypen sind einschlägig und haben ihre Berechtigung in einem Fragebogen; die halboffene Frage hingegen sollte möglichst vermieden werden. Denn bei dieser konnte sich der Fragensteller zumeist nicht zu einer geschlossenen Frage durchringen; stattdessen wird eine geschlossene Frage als (meistens letzte) Antwortmöglichkeit noch mit einem Zusatz „Sonstiges, bitte nennen ...“ garniert.

Kognitionspsychologische Forschung hat gezeigt, dass die Entscheidung für eine offene oder geschlossene Frage erheblichen Einfluss auf das (Antwort-)Ergebnis hat. Ein Befragter kann sich nicht davor schützen, eine ihm gestellte Frage auf den Zweck hin zu interpretieren. Und die beiden Fragetypen – offen oder geschlossen – können zu unterschiedlichen Interpretationen und damit zu unterschiedlichem Antwortverhalten führen.⁷

Bei der Wahl zwischen offener und geschlossener Frage kann grundsätzlich als Entscheidungshilfe der Aspekt der Vergleichbarkeit herangezogen werden: Ist Vergleichbarkeit notwendig oder wünschenswert? Dann bieten sich geschlossene Fragen eher an; auch eine höhere Durchführungs- und Auswertungsobjektivität und eine leichtere Beantwortbarkeit sowie ein geringerer Aufwand bei der Auswertung ist dann erwartbar. Offene Fragen bieten sich hingegen an, wenn die erwartbaren Antwortkategorien noch zu wenig bekannt oder gänzlich unbekannt sind.⁸

Zuletzt: Es hält sich hartnäckig das Vorurteil, offene Fragen seien per se leichter zu formulieren als geschlossene Fragen. Das mag in einigen Fällen tendenziell so sein; doch gerade bei der Formulierung von offenen Fragen (siehe auch Konstruktionstipp 3) ist mit besonderer Sorgfalt vorzugehen.



FAZIT: Vermeide halboffene Fragen und entscheide Dich gezielt für eine offene oder geschlossene Frage (mit Einfach- oder Mehrfachnennung). Beachte dabei unterschiedliche Interpretationstypen und lasse auch bei offenen Fragen höchste Sorgfalt walten.

Für geschlossene Fragen – insbesondere mit Einfachnennung – gibt es noch Weiteres zu bedenken.

KONSTRUKTIONSTIPP 5:

Welche Skala ist geeignet?

Für geschlossene Fragen mit Einfachnennung lassen sich bei den Antwortkategorien nach der klassischen Klassifizierung von Stevens Nominal-, Ordinal-, Intervall- und Ratio-Skalen unterscheiden.⁹

Unterscheidung der Skalenniveaus

Um die geeignete Antwortskala zu ermitteln, hilft es, sich deren Charakteristika noch einmal kurz in Erinnerung zu rufen. So ist die typische „Sonntagsfrage“ nach der Wahlabsicht in der Bevölkerung eine repräsentative, nominal skalierte Fragestellung:

Wenn am nächsten Sonntag Bundestagswahlen wären, welche Partei würden Sie dann mit Ihrer Zweitstimme wählen?

CDU/CSU SPD Bd. 90/Die Grünen FDP AfD Die Linke Andere

Bei ordinal skalierten Fragen stehen die Antwortmöglichkeiten in einer relationalen Beziehung zueinander – sie unterliegen einer Rangordnung. Wichtig hierbei ist, dass die Gleichabständigkeit der Antwortkategorien nicht gegeben sein muss. Beispiel:

Wie stark interessieren Sie sich für Politik:

sehr stark stark mittel wenig überhaupt nicht?

Bei Fragen, die auf einer Intervall-Skala beantwortet werden, sind die Abstände zwischen den Skalenpunkten der Antwort definitionsgemäß gleich. Wenn A, B, C und D einander folgen, dann ist der Abstand von B zu A gleich dem Abstand von D zu C. Solche Skalen sind im Rahmen von Befragungen nach Weiterbildungsmaßnahmen eigentlich eher selten anzutreffen. Da der Mittelwert (genauer: das arithmetische Mittel) in der Praxis vielfach jedoch beispielsweise aus Bewertungen auf einer Skala von 1 bis 10 gebildet wird, müssten solche Skalen aber eigentlich vorliegen; denn diese mathematischen Operationen setzen „eigentlich“ intervallskalierte Variablen voraus. Die Lösung ist: Man nutzt endpunktbenannte Skalen und tut so, als seien die Skalenpunkte gleichabständig.

trifft überhaupt nicht zu	○	○	○	○	○	○	○	trifft voll und ganz zu
---------------------------	---	---	---	---	---	---	---	-------------------------

Ratio-Skalen sind im Grunde nichts anderes als Intervall-Skalen mit einem echten Skalennullpunkt.

sehr unzufrieden	-3	-2	-1	0	1	2	3	sehr zufrieden
------------------	----	----	----	---	---	---	---	----------------

Die theoretische Diskussion um die „einzig richtige“ Skala für einen Fragebogen ist so alt wie die moderne Umfrageforschung selbst. Aus praktischer Sicht ist eine Intervall-Skala mit künstlichem Skalennullpunkt wohl das für geschlossene Fragen geeignetste Skalenniveau.

Gerade oder ungerade Skalen

Hat man ein Skalenniveau definiert, stellt sich als nächstes die Frage: Eine gerade oder eine ungerade Anzahl von Antwortmöglichkeiten verwenden? Schon aus der oben angeführten Forderung nach einem Skalennullpunkt ergibt sich folgerichtig: eine ungerade Anzahl. Hierfür sprechen auch noch andere Gründe:

Nutzt man in einem Fragebogen eine geradzahlige Skala, so zwingt man die Befragten zu einer Entscheidung. Es handelt sich dann um ein „Forced-Choice-Format“. Ausgeschlossen ist damit für die Befragten die Chance, sich bewusst und gezielt für die mittlere Position zu entscheiden – obwohl sie sich vielleicht inhaltlich dort zugehörig fühlen. Dieses kann (im schlechtesten Fall) dazu führen, dass der Befragte „irgendeine“ Antwort gibt.¹⁰ Daher sollte der Skalennullpunkt angeboten werden – denn er existiert in der Realität meistens ebenso.

Es sollte ferner eine außerhalb der Skala befindliche Restkategorie (z.B. „Kann ich nicht beurteilen“) angeboten werden. Diese „Fluchtkategorie“ mindert die Anzahl der Nennungen auf dem mittleren Skalenwert bei einer ungeraden Skalenzahl und sorgt so für ein realistisch(er)es Bild.¹¹

Wahl der Skalenbreite

Die Skalenbreite kann grundsätzlich von der vermuteten Abstraktionsfähigkeit der Befragten abhängig gemacht werden. Aber Skalen mit 10 oder sogar mehr Skalenpunkten sollten mindestens mit Skepsis betrachtet, wenn nicht gar prinzipiell abgelehnt werden.

Würden Sie diese Veranstaltung weiterempfehlen?

| 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 |

Denn den Unterschied zwischen der Beantwortung mit „7“ und der mit „8“ kann zumeist weder der Befragte noch der Fragensteller exakt benennen. So kann eine zu große Anzahl an Skalenpunkten a) zu einer Überforderung der Befragten führen oder b) im Rahmen der Auswertung zu unzulässigen Schlussfolgerungen. Bei mehr als sieben Skalenpunkten kann von den Befragten praktisch kaum noch eine Differenzierung zwischen zwei Skalenwerten vorgenommen werden. Dies hat zumeist die Konsequenz, dass wiederum die extremen Skalenwerte häufiger besetzt sind.

Bei weniger als fünf Skalenpunkten hingegen wird den Befragten zu wenig Spielraum für eine hinreichend differenzierte Bewertung geboten, woraus sich ein Widerwille gegen die Beantwortung entwickeln kann. So besteht beispielsweise eine dreistufige Skala praktisch nur aus extremen Positionen.

Empfehlenswert ist eine 7er-Skala. Bei dieser ist es noch möglich, jedes Skalenelement sprachlich zu benennen (und damit zu unterscheiden). Zum Zeitpunkt der späteren Auswertung sollte diese 7er-Skala dann intern auf eine 5er-Skala reduziert werden. Dazu werden die Extremantworten in der Auswertung um eins zur Mitte hin verschoben. Damit werden zum einen extreme Positionen zusammengefasst und der Neigung mancher Menschen, Extreme nicht zu nutzen, wird entgegengearbeitet.

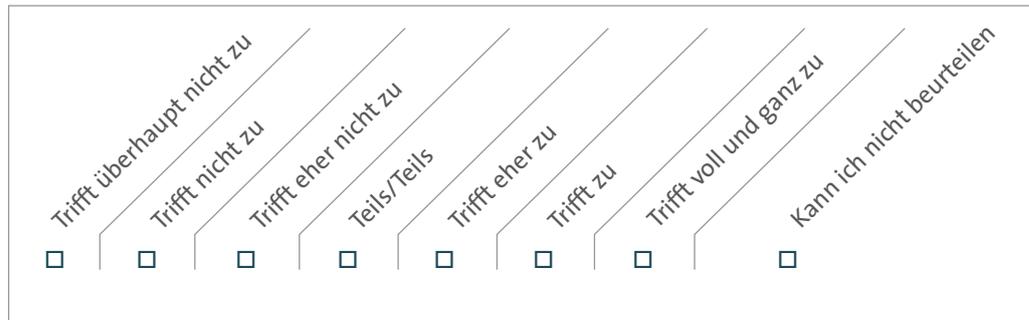
Wichtig ist ferner: Die Antwortkategorien einer Frage müssen disjunkt (überschneidungsfrei) sein.

Wachsen oder fallen: Richtung der Skalen

Zuletzt gilt es noch, die Richtung der gewählten Skala zu betrachten. Wird der höchste/beste Skalenwert (z.B. „sehr zufrieden“) am linken oder am rechten Skalenendpunkt notiert? Für unseren Kulturraum gilt: Optisch präsentierte Skalen werden am besten „von links nach rechts“ notiert. D.h.: „überhaupt nicht wichtig“ wird am linken und „sehr wichtig“ am rechten Skalenende notiert. Eine umgekehrte Repräsentation wäre kontraintuitiv.

Im Sinne einer Äquidistanz ist es dann noch wichtig, dass die beiden Pole am äußeren Rand der Skala dieselbe Aussage nur in unterschiedlicher Richtung wiedergeben. So sollten bei der Beschreibung des linken (=niedrigsten) Skalenwertes negative Begriffe und Zeichen („5“ oder „lehne voll und ganz ab“) vermieden werden (stattdessen: „stimme überhaupt nicht zu“).

Im Ergebnis könnte eine mustergültige Skalierung dann wie folgt aussehen:



FAZIT: Wähle eine 7er-Intervall-Skala mit disjunkten Antwortkategorien, einem echten Skalenmittelpunkt sowie einer Fluchtkategorie; wobei die Richtung von links (schlechtster Wert) nach rechts (bester Wert) orientiert ist.

KONSTRUKTIONSTIPP 6:

Passt die Skala zur Fragestellung?

Innerhalb eines Fragebogens kommt – sofern irgend möglich – am besten nur eine Skalenart zum Einsatz. Dazu ist es notwendig: a) eine Skala zu definieren, die zu allen im Fragebogen vorkommenden Fragen passt; b) zur Skala passende Fragen zu formulieren.

Passen Frage und Skala nicht zueinander, entsteht für den Befragten Interpretationsspielraum und es wird möglicherweise versucht, die Absicht hinter der Frage zu antizipieren und eine sozial erwünschte Antwort (siehe Auswertungstipp 1) zu geben. Beispiel:

Für wie empfehlenswert halten Sie <NameDesSeminars> im Business-Umfeld?
| 10% | 30% | 50% | 70% | 90% |

Hier sind nicht nur die Skalenelemente als Prozentwerte unerwartet. Auch die fehlenden Pole (0% und 100%) machen misstrauisch.

Ferner gilt es, darauf zu achten, dass die Wertung ausschließlich über das Skalenniveau in den Antworten erfolgt. Die Frage selbst sollte keine Wertung enthalten.



FAZIT: Nutze nur eine Skalenart im Fragebogen und stelle eine Passung zwischen Antwortelementen und Fragestellung sicher.

KONSTRUKTIONSTIPP 7:

Orientierung am Befragten!

Der Blick auf den „Happy sheet“ ist für die Befragten zumeist eine der letzten Aktionen bei der Teilnahme an einer Weiterbildung; zum Abschluss einer womöglich anstrengenden Maßnahme sehen sie sich konfrontiert mit einem ihnen vielleicht gänzlich neuen Fragenkatalog. Dieser sollte ihnen auf keinen Fall den Eindruck eines undurchsichtigen Fragen-Ungetüms vermitteln, wie vielen durch so manches Formular von Behörden oder Ämtern bekannt.

Wichtig ist deshalb: Aufmachung, Gestaltung des Fragebogens, Reihenfolge der Fragen etc. dürfen sich nicht etwa daran orientieren, dass der Fragebogen später einfach auswertbar ist oder Ähnliches; entscheidendes Kriterium muss vielmehr die prompte Verständlichkeit auf Seiten der Befragten sein. Die Reihenfolge der Fragen sollte nicht willkürlich anmuten, sondern eine inhärente sinnvolle Struktur auch einem Laien sogleich ersichtlich sein.

Auch ist wichtig: Als Sparringspartner für die Weiterentwicklung der Bildungsmaßnahme, des Vortrags etc. eignen die Befragten sich nicht! Sie verfügen üblicherweise nicht über das Experten-Know-how, um Fragen zu beantworten wie etwa:

Wie beurteilen Sie die Vollständigkeit der fachlichen Inhalte?

Eine solche Frage setzt fundiertes Fachwissen voraus, das eigentlich auf Seiten des Weiterbildungsanbieters bzw. Referenten bzw. Vortragenden vorhanden sein sollte (der durch eine solche Frage seine eigene thematische Unsicherheit zum Ausdruck bringt).

Hilfreich ist es, den Fragebogen vor einem Echteinsatz einem so genannten Pretest zu unterziehen: ihn insgesamt oder auch einzelne Fragen Personen testweise vorzulegen. Sofern sich Verständnisschwierigkeiten ergeben, ist nicht die Testperson zu blöd, sondern der Fragebogen ist noch nicht „reif“.



FAZIT: Strukturiere und gestalte den Fragebogen so, dass er für mit den Inhalten nicht Vertraute leicht verständlich und handhabbar ist, und führe einen Pretest durch.

Auswertung von Fragebögen: Unsinn mit statistischer Genauigkeit

„Ich glaube keiner Statistik, die ich nicht selbst gefälscht habe“ – eine viel zitierte Aussage von Winston Churchill. Eine, die wohl auch auf die Auswertung von so mancher Seminar- oder Veranstaltungsevaluation zu passen scheint. Doch Vorsicht: Auch hier liegt ein verbreiteter Irrtum vor. Denn dieses Zitat stammt wohl nicht von Churchill. Urheber, historischer und sonstiger Kontext dieses Zitats gelten bis heute als ungeklärt.

Auch bei der Auswertung von Feedbackbögen besteht die Gefahr methodischer Fehler. Insbesondere statistische Fehler liegen hier nahe.

AUSWERTUNGSTIPP 1:

Hast Du die Toten berücksichtigt?

Immer wieder erstaunt es, zu welchen Traumergebnissen manche Trainer und/oder Speaker in ihren Teilnehmerfeedbacks kommen. 100 Prozent positives Feedback oder zusammengefasstes Teilnehmerfeedback einer Veranstaltung mit über 600 Teilnehmenden: Schulnote 1,0. Wow! Aber: Kann das wirklich sein?

Eine erste Ursache solcher Traumergebnisse kann darin liegen, dass die Teilnehmenden direkt im Anschluss an eine Veranstaltung befragt werden. Zu diesem Zeitpunkt haben aber diejenigen, denen es womöglich nicht gefallen hat, die Veranstaltung vielleicht bereits verlassen. Oder aber es werden in der Berechnung nur diejenigen einbezogen, die auch einen Feedbackbogen retournieren. Doch: Was ist mit dem Rest?

Wenn wir ehrlich rechnen und uns nicht nur von Spitzenwerten blenden lassen, dann werden Extremwerte wie „100 Prozent“ die absolute Ausnahme bilden – und das ist auch gar nicht schlimm.



FAZIT: Berücksichtige in der Feedbackbogen-Auswertung auch die nicht abgegebenen Feedbackbögen als Einflussgrößen.

Gibt es ein wirksames Gegenmittel – lässt sich die so genannte „Sterberate“ in einer Evaluation eliminieren? Etwa, indem jeder Teilnehmende (natürlich liebenswürdig) verpflichtet wird, einen Feedbackbogen auszufüllen? Diese Überlegung bringt uns zum zweiten Tipp.

AUSWERTUNGSTIPP 2:

Sind die Antworten vielleicht sozial erwünscht?

Befragte neigen dazu, ihr Antwortverhalten an dem zu orientieren, was sie für sozial erwünscht halten bzw. wovon sie annehmen, es werde von ihnen erwartet. Insofern gilt es beispielsweise stets zu fragen, welche negativen Folgen, Sanktionen etc. die Befragten zum Zeitpunkt der Befragung möglicherweise erwarteten, was dann ihr Urteil beeinflusst haben könnte.

Die soziale Norm – gesetzt etwa durch ein beauftragendes Unternehmen oder durch ein Veranstaltungssetting – kann die nach Auffassung des Befragten erwünschte Skalenbreite für eine Beantwortung – unabhängig von der tatsächlichen Situation – künstlich verengen. Dann braucht es für schlechte (aber vielleicht ehrliche) Bewertungen viel Selbstsicherheit.

Ebenso kann eine mehr oder minder unabsichtliche Neigung zur positiven Selbstdarstellung Einfluss auf die Beantwortung der Fragestellungen haben und ehrliche Antworten verhindern.

Gleiches kann auch erreicht werden durch die Erzeugung von Druck mittels Formulierung von (Folge-)Fragen:

Wenn Ihre Bewertung niedriger als 10 ist, was könnten wir konkret verbessern, um es für Sie zu einer 10 zu machen?

Nicht jeder Befragte verfügt über das notwendige Selbstvertrauen, um sich einem solchen (wenn auch nur gefühlten) sozialen Druck entgegenzustellen.



FAZIT: Überprüfe, inwieweit die Befragten frei in der Beantwortung des Feedbackbogens waren, und berücksichtige dies in der Ergebnisinterpretation.

AUSWERTUNGSTIPP 3:

Median statt/und Mittelwert

In vielen Fragebögen werden die Antworten entweder von vorne herein nach Schulnoten oder anderen numerischen Antwortoptionen erfasst oder aber Skalenwerte werden zu Auswertungszwecken im Nachgang der Befragung in numerische Werte konvertiert. Ziel ist es, für die Auswertung ein zusammenfassendes Ergebnis zu berechnen. Hierzu gibt es dann verschiedene mathematische Methoden und Variablen, die berechnet und für eine sorgfältige Auswertung in die Interpretation einbezogen werden.

gen werden können. Auf die Vielfalt dieser Operationen kann ich jetzt und hier nicht eingehen, also betrachte ich an dieser Stelle nur die häufigste Berechnungsform: die Zusammenfassung der Befragungsergebnisse über den Mittelwert.

Bei numerischen Antworten stellt für die meisten Auswertenden die Berechnung des Mittelwerts sowieso die einzige (bekannte) Lösung dar. Doch auch hier gilt: Was richtig klingt, muss noch lange nicht richtig sein. Denn der Mittelwert hat mathematisch eine entscheidende Schwäche: Er überbewertet Extremwerte in der Skalierung. Anders verhält es sich mit dem Median.

Ein Beispiel: Ein Seminar ist völlig „in die Hose gegangen“. Es werden nur fünf Feedbackbögen retourniert – was wir gemäß Auswertungstipp 1 berücksichtigen sollten, an dieser Stelle aber erst einmal ignorieren. Von diesen fünf Feedbackbögen hat das Seminar von vier Befragten im Schulnotensystem eine glatte Sechs erhalten; ein Teilnehmer hat das Seminar allerdings (durchaus überraschend) mit einer Eins bewertet. Würde nun der Mittelwert gebildet, so wäre das Gesamtergebnis eine Fünf – damit zwar immer noch die Note „mangelhaft“, aber immerhin. Wird hingegen der Median berechnet, kommt das Ergebnis der Realität wohl viel näher: Sechs, d.h. „ungenügend“.

Der Median ist der Wert, der bei den nach Antworten sortierten Rückläufen genau in der Mitte liegt. Der Median teilt die Stichprobe dazu in genau zwei Hälften, so dass die eine Hälfte Werte aufweist, die kleiner als der Median sind, und die andere Hälfte Werte, die größer sind. Daher wird der Median von Extremwerten (Ausreißern) praktisch kaum beeinflusst und ist so eine robustere Schätzung der zentralen Tendenz.

Deshalb sollte bei der numerischen Zusammenfassung von Fragebögen und der Auswertung des Gesamtfeedbacks einer Befragung entweder nur der Median (anstelle des Mittelwerts) berechnet werden oder zumindest beide Werte.



FAZIT: Berücksichtige, dass es eine bessere Alternative zum Mittelwert gibt.

AUSWERTUNGSTIPP 4:

Erfolgszuschreibungen mit maximaler Genauigkeit

Lernerfolg und Kompetenzzuwachs kann viele Ursachen haben – nicht immer ist beispielsweise der Referent der dafür Verantwortliche; ebenso besteht die Möglichkeit, dass der Austausch mit anderen Teilnehmern – oder auch Geschehnisse außerhalb des Veranstaltungssettings – zu dieser positiven Entwicklung geführt haben. Das ist bei der Abfrage des Lernerfolgs und Kompetenzzuwachses und Ähnlichem im Auge zu behalten.

So sollte bei der Interpretation von Antworten stets kritisch geprüft werden: Ist die festgestellte Änderung bei Teilnehmenden tatsächlich auf die evaluierte Veranstaltung und den Beitrag des Referenten oder Vortragenden zurückzuführen? Ansonsten droht an diesem Punkt die Gefahr einer unberechtigten Zurechnung positiver Ergebnisse für den Referenten. Durch geeignete Formulierung der Fragen kann hier einer Fehleinschätzung entgegengewirkt werden.



FAZIT: Um unberechtigte Effektzuschreibungen auszuschließen, Sorge dafür, dass Fragen nur direkte Einflüsse des Seminars/Vortrags auf die Befragten messen.

AUSWERTUNGSTIPP 5:

War nicht völlig egal, was gefragt wurde?

Tendenzen, Fragen unabhängig von ihrem Inhalt zu beantworten, sind die so genannten Response-Set-Tendenzen. Es kann beispielsweise bei Befragten die Zustimmungstendenz bestehen; damit ist die seitens empirischer Forschung beschriebene Neigung gemeint, einer – entsprechend suggestiv formulierten – Frage eher zuzustimmen als sie zu verneinen.

In schriftlichen Befragungen gibt es weiter den Primacy-Effekt: die Tendenz zur Auswahl einer der ersten Antworten einer Liste. Befragte wählen dann schlicht die erste Antwort, die sie befriedigt, ungeachtet weiterer Antwortalternativen. Diese Tendenz kann insbesondere durch fehlendes Wissen oder mangelndes Interesse am Befragungsgegenstand verstärkt werden. Vergleichbar ist auch das Vorgehen, beim Ankreuzen ohne langes Nachdenken und unabhängig vom Inhalt der Fragen stets einem bestimmten Muster zu folgen oder immer den (nahezu) gleichen Bereich einer Skala anzukreuzen. Das Risiko dieser Vorgehensweise nimmt mit der Anzahl der Fragen deutlich zu.¹²



FAZIT: Achte auf so genannte „Response-Set-Tendenzen“ bereits bei der Entwicklung des Fragebogens und beschränke die Anzahl der Fragen auf ein Minimum. Achte beim Auswerten der Antworten auf Musterbildungen; berücksichtige diese ggf. in der Ergebnisinterpretation.

Fazit

Ein wirklich guter, nützlicher Fragebogen ist nicht im Nu „gestrickt“; es handelt sich dabei nicht um eine klassische Aufgabe für einen studentischen Praktikanten oder etwas, das „on-the-fly“ erledigt wird. Vielmehr erfordert ein guter Feedbackbogen Denkarbeit – sowohl bei der Konstruktion als auch bei der Auswertung. Nicht zuletzt deshalb gibt es Experten, die sich eingehend mit diesem Thema auseinandersetzen und Unterstützung bieten.

Den perfekten Feedbackbogen gibt es vielleicht nicht – aber man kann ihm zumindest nahe kommen. Mit einem Fragebogen, der die Lebenszeit derer wert ist, die ihn konzipieren, die ihn beantworten oder ihn auswerten.

Anmerkungen

- 1 Beispielsweise Russell (1997)
- 2 Siehe dazu Russell (1997)
- 3 Die Beispiele aus der Feedbackbogen-Praxis wurden freundlicherweise von Mitgliedern der German Speakers Association e.V. (GSA) – der Vereinigung der deutschsprachigen professionellen Redner – als Anschauungsmaterial für diesen Artikel zur Verfügung gestellt. Ein herzliches Dankeschön an Michael Rossié für die Sammlung dieser Fragebögen von den Mitgliedern!
- 4 Becker (2010), S. 2
- 5 Büning (2012): S. 85ff
- 6 Coltart, Nina (1991), S. 447
- 7 Porst (2014), S. 61ff
- 8 Diekmann (2009), S. 476ff
- 9 Stevens (1946)
- 10 Porst (2014), S. 84
- 11 Porst (2014), S. 84
- 12 Porst (2014), S. 17

Literaturverzeichnis

- Becker, Fred G. / Wittke, Ingrid / Friske, Vanessa (2010):** „Happy Sheets“. *Empirische Befragung von Bildungsträgern*. Diskussionspapier / Fakultät für Wirtschaftswissenschaften, Bielefeld: Universität Bielefeld, Fak. für Wirtschaftswissenschaften.
- Büning, Norbert (2012):** *Lernen im Unternehmen*, Redline, München
- Coltart, Nina (1991):** *The silent patient*. In: *Psychoanalytic Dialogues*, Vol. I - No. 4, S. 439-453
- Diekmann, Andreas (2009):** *Empirische Sozialforschung*, Rowohlt, Reinbek
- Porst, Rolf (2014):** *Fragebogen*, Springer, Wiesbaden
- Russell, Jeffrey Burton (1997):** *Inventing the Flat Earth: Columbus and Modern Historians*, Westport Conn., London.
- Stevens, Stanley Smith (1946):** *On the Theory of Scales of Measurement*. In: *Science*, Vol. 103 - No. 2684, S. 677–680



Über den Autor

Ingo Radermacher, Geschäftsführer der radermacher-consulting GmbH, berät vor allem mittelständische Unternehmen im strategischen IT-Einsatz und bietet Impulsvorträge an.

Als **Keynote-Speaker** macht sich Ingo Radermacher das informatische und das mathematische Denken zunutze, um Menschen außerhalb der IT verblüffende Impulse zum Selberdenken zu geben.

Sein Thema ist: **Klarheit!** Durch eine Rückkehr zum klaren Denken, Reden und Handeln verhilft er zu klugen Problemlösungen und Entscheidungen. So wird ein intelligenter Umgang mit Informationen möglich.

Ingo Radermacher
Der Entscheidungsphilosoph®

Dipl.-Wirt.-Inf. (FH)

Ingo Radermacher

Keynote Speaker

Professional Speaker GSA (SHB)

Certified Management Consultant (CMC/BDU)

Grünenborn 1 | 53797 Lohmar

Tel.: 02206 . 917 91-0

Fax: 02206 . 917 91-60

kontakt@ingoradermacher.de

www.ingoradermacher.de

Der Entscheidungsphilosoph® ist eine Marke
der radermacher-consulting GmbH
www.radermacher-consulting.de