

Fragebogen

Mithilfe geeigneter Fragen subjektive Ausstellungserfahrungen erforschen

Eva Specker und Helmut Leder

Einleitung

Fragebögen sind eine klassische Untersuchungsmethode in der Psychologie, aber auch in den Museumswissenschaften und anderen (verwandten) Disziplinen. Der wesentliche Vorteil von Fragebögen liegt darin, dass sie eine direkte Erfassung subjektiver Erfahrungen ermöglichen, die mit anderen Methoden oft schwieriger zu erheben wären. Wenn wir beispielsweise wissen möchten, ob Museumsbesucher:innen ein Kunstwerk mögen, erscheint es eine naheliegende und valide Messmethode, sie direkt zu fragen, ob ihnen das Kunstwerk gefällt (vorausgesetzt, die Antwort ist ehrlich). Das bedeutet nicht, dass keine anderen Methoden verfügbar wären. So könnte man beispielsweise auch die Betrachtungszeit messen (Leder, Mitrovic und Goller 2016) oder die Pupillengröße (Kuchinke et al. 2009), die ebenfalls mit Gefallen assoziiert werden. Jedoch können viele andere Faktoren (neben dem Gefallen) die Betrachtungszeit beeinflussen, sodass dies nur eine indirekte Messung darstellt. Dies gilt für viele physiologische Messungen. Darüber hinaus können Fragebögen differenziertere Informationen über die Erfahrungen der Befragten liefern als andere Methoden. So können wir zwar positive und negative emotionale Reaktionen messen, indem wir Gesichtsmuskeln, Lächeln oder Stirnrunzeln analysieren (Gerger et al. 2011). Ein Fragebogen erlaubt es jedoch, eine Vielzahl verschiedener emotionaler Reaktionen zu unterscheiden (Schindler et al. 2017) und zu bestimmen. Insbesondere, wenn wir ästhetische Erfahrungen untersuchen möchten, bieten Fragebögen in der Regel die direkteste Möglichkeit, diese differenziert zu erfassen. Geht es speziell um die Anwendung in einem Ausstellungskontext, haben Fragebögen zudem den Vorteil, dass sie kostengünstig (es sind keine teuren Geräte erforderlich), zeiteffizient und relativ einfach umzusetzen sind.

In mancherlei Hinsicht können Fragebögen jedoch darin täuschen, wie einfach sie zu erstellen sind (z. B. Simms 2008). Gerade da wir alle im Alltag Sprache verwenden, mag es sehr einfach erscheinen, einen Fragebogen zu erstellen, indem man

einfach ein paar passende Fragen eintippt. In diesem Kapitel erklären wir, warum es sich lohnt, Zeit in die Entwicklung von Fragebögen zu investieren, um eine passende Messung sicherzustellen. Wir beschränken uns hier auf quantitative Messungen. Zwar sind Fragebögen flexibel und könnten somit auch offene Fragen enthalten, die qualitative Analysen ermöglichen. Fragebögen lassen sich auch in Form von Tests verwenden, bei denen richtige oder falsche Antworten möglich sind – beispielsweise bei der Überprüfung von Kunstwissen (z. B. Specker 2021, Specker, Cotter und Kim 2023, Specker, Forster et al. 2020). Dies ist eine durchaus häufig eingesetzte Methode, doch widmen wir uns im vorliegenden Kapitel der quantitativen Erfassung subjektiver Erfahrungen (im weitesten Sinne) von Ausstellungen.

Auch konzentrieren wir uns hauptsächlich auf die Entwicklung von Items (d. h. die Erstellung eigener Fragen) und weniger auf andere Aspekte der Skalen- oder Fragebogenentwicklung (siehe z. B. Gehlbach und Brinkworth 2011 für einen umfassenderen Überblick). Dieser Fokus wurde gewählt, um Forscher:innen zu befähigen, Items/Fragen zu entwickeln, die den spezifischen Bedürfnissen ihrer jeweiligen Studie entsprechen, nicht jedoch, ein tiefgreifendes Wissen über die Entwicklung von Skalen/Fragebögen zu vermitteln. Wichtig zu beachten ist, dass für Einzelitems (z. B. ‚Wie sehr gefällt Ihnen dieses Kunstwerk?‘) keine statistischen Validitätsnachweise erbracht werden können. Die Bewertung solcher Einzelitems stützt sich daher hauptsächlich auf die sogenannte ‚Augenscheinvalidität‘ (d. h., ob die Frage zu messen scheint, was gemessen werden soll). Dennoch kann die Validität solcher Einzelitems durch die Formulierung der Fragen verbessert werden – daher liegt unser Fokus auf der Entwicklung von Items.

Ziele der Methode

In den meisten Fällen werden Fragebögen zur Messung der abhängigen Variablen verwendet. Ein entscheidender Punkt ist dabei, dass in (quasi-)experimentellen Designs eine unabhängige Variable (oder ‚Ursache‘) und eine abhängige Variable (oder ‚Wirkung‘) vorhanden sind. Allerdings können Fragebögen auch verwendet werden, um andere abhängige Variablen oder unabhängige Variablen zu messen. In allen Fällen ist es entscheidend, dass die Messung valide ist – das heißt, dass tatsächlich gemessen wird, was gemessen werden soll. Dies ist oft leichter gesagt als getan, da die Validierung eines Fragebogens in der Regel viel Zeit und Mühe erfordert (z. B. Specker 2021, Specker, Cotter und Kim 2023, Specker, Forster et al. 2020).

Schritt-für-Schritt-Anleitung

Wenn man einen Fragebogen verwenden will, sollten man zunächst prüfen, ob das theoretische Konstrukt mit einem Fragebogen überhaupt messbar ist (siehe Schritt 1 unten), und natürlich auch, ob es bereits einen Fragebogen gibt, den man nutzen könnte. Ist dies der Fall, kann man die Schritte 2–4 befolgen, um den Fragebogen zu erstellen.

Schritt 1: Definieren Sie zunächst das theoretische Konstrukt, das Sie messen möchten (Gehlbach und Brinkworth 2011). Was genau wollen Sie messen? Gefallen? Schönheit? Stress? Nachdem dies entschieden ist, können Sie in der Literatur nachsehen, ob es validierte Skalen gibt. Die Verwendung einer validierten Skala stellt sicher, dass Ihre Messung valide ist, auch lässt sich dann Ihre Arbeit besser mit anderen Studien vergleichen. Gibt es eine solche Skala nicht, können Sie untersuchen, wie andere dieses Konstrukt in der Vergangenheit gemessen haben. Warum das Rad neu erfinden (Clark und Watson 1995)? Trotzdem kann es sein, dass Sie nichts Passendes finden, z. B. weil Sie etwas Spezifisches zu einer Ausstellung oder einem Museum abfragen möchten.

Somit wenden wir uns nun der Entwicklung von Items zu. Einige Empfehlungen mögen im Grunde selbstverständlich erscheinen, sind aber in der Praxis eher schwierig umzusetzen. Das vorgeschlagene Vorgehen ist dabei häufig schon empirisch erprobt und basiert auf Best-Practice-Empfehlungen (siehe z. B. Gehlbach und Brinkworth 2011, Simms 2008 für kurze Übersichten). Wir haben uns für eine Beschreibung in leicht verständlicher Sprache ohne Fachjargon entschieden, an Beispielen aus dem Bereich der Museums- und Ausstellungswissenschaften.

Schritt 2: Hat man eine Entscheidung darüber getroffen, was man messen möchte, gilt es das Wie festzulegen. Beim Erstellen eigener Fragen muss man sich zuallererst über das Format Gedanken machen. Wählt man hier eine feststellende Aussage, z. B.: ‚Wie sehr gefällt Ihnen dieses Kunstwerk?‘ oder: ‚Inwieweit stimmen Sie folgender Aussage zu: ‚Mir gefällt dieses Kunstwerk?‘ Zu beachten ist, dass es grundsätzlich ratsam ist, das Antwortformat ‚stimme zu/stimme nicht zu‘ zu vermeiden (z. B. Krosnick 1999). Der Grund dafür ist, dass dieses Format eine kognitiv anspruchsvolle Aufgabe darstellt, die Fehler begünstigen und den Aufwand für die Teilnehmenden unnötig vergrößern kann (Gehlbach und Brinkworth 2011), dies gilt insbesondere bei invers formulierten Items oder doppelten Verneinungen (Swain, Weathers und Niedrich 2008). Aus demselben Grund ist es ratsam, bei mehreren Fragen zum gleichen Konstrukt ein einheitliches Frageformat beizubehalten. Wenn Sie wechseln, also beispielsweise für Kunstwerk 1 eine Frage und für Kunstwerk 2 eine feststellende Aussage verwenden, kann dies die Teilnehmenden Ihrer Studie

verwirren. Teilnehmende werden die Skala, die Sie ihnen vorlegen, einfacher und schneller (genau) ausfüllen können, wenn die Struktur der Fragen konsistent bleibt.

Aus diesem Grund ist es auch ratsam, das gleiche Antwortmuster beizubehalten und die Antwortmöglichkeiten konsistent zu formulieren. Wenn Sie beispielsweise eine 7-Punkte-Skala verwenden, sollte 1 dann immer ‚gar nicht‘ und 7 immer ‚sehr viel‘ bedeuten. Manchmal werden Änderungen im Format vorgenommen, „um die Aufmerksamkeit der Teilnehmenden zu stimulieren“ oder „dafür zu sorgen, dass sie ehrlich antworten“ (Gehlbach und Brinkworth 2011). In der Praxis ist es jedoch so, dass eine Änderung der Skalenstruktur, z. B. eine Umkehrung der Bewertung, die Zuverlässigkeit (z. B. Benson und Hocevar 1985) beeinträchtigt und zu „Fehlreaktionen“ führt (Swain, Weathers und Niedrich 2008). Beachten Sie, dass dies sowohl für die Formulierung der Items als auch für die Art der Antwortformulierung gilt. Konkret sollten die Items so gestaltet sein, dass eine hohe Bewertung (z. B. ‚sehr viel‘) stets eine positive Reaktion ausdrückt (z. B. dass einem eine Ausstellung gefällt) und sich die Bedeutung nicht verändert.

Schritt 3: Beim Formulieren Ihrer Fragen sollte Ihr Hauptziel Klarheit sein (Clark und Watson 1995, Gehlbach und Brinkworth 2011, Simms 2008). Dies ist jedoch leichter gesagt als getan. Wir empfehlen:

1. Jede Frage sollte sich nur auf einen zentralen Gedanken beziehen.
2. Seien Sie präzise.
3. Formulieren Sie kurz.
4. Vermeiden Sie komplizierte Formulierungen oder unklare Konstrukte.
5. Lassen Sie unwichtige Informationen weg.
6. Verwenden Sie positive Sprache und vermeiden Sie doppelte Verneinungen.
7. Vermeiden Sie Begriffe wie ‚alle‘ und ‚keine‘.
8. Vermeiden Sie auch vage Begriffe wie ‚häufig‘ oder ‚manchmal‘.

Viele dieser Regeln leiten sich daraus ab, wie Menschen Sprache verstehen. So sind etwa doppelte Verneinungen schwierig zu begreifen (u. a. Benson und Hocevar 1985). Anders gesagt: Unklare Fragen erschweren es den Teilnehmenden, präzise zu antworten, und Ihnen als Forschenden, die daraus gewonnenen Daten zu interpretieren. Darum geht es im Wesentlichen in Punkt 7 und 8. Zum Beispiel ist ‚Ich habe mich vor allem für die Gemälde interessiert‘ ein klassisches Beispiel für eine doppelläufige Frage, also eine Frage, die mehr als eine Sache bewertet (d. h., sie verstößt gegen Empfehlung 1). Ein weiteres Beispiel wäre die Abfrage ‚Meine Museumserfahrung war informativ und angenehm‘. Der/die Teilnehmende könnte unschlüssig sein, wie darauf zu antworten ist, wenn die Erfahrung nur informativ oder nur angenehm, aber nicht beides war.

Hier kommt es also auf die Intention der Frage an: Was möchten Sie messen? (siehe auch Schritt 1). Wenn Sie beispielsweise wissen möchten, ob die Teilnehmenden relativ mehr an Gemälden oder an Skulpturen interessiert sind, könnte es sinnvoll sein, dies direkt zu fragen, z. B.: „Waren Sie mehr an Gemälden oder an Skulpturen interessiert?“ (wie Reitstätter et al. 2020) mit beispielsweise Skalenpunkten von 1 „mehr an Gemälden“ bis 7 „mehr an Skulpturen“. Wenn Sie wissen möchten, wie interessiert die Teilnehmenden individuell an Gemälden und Skulpturen waren, dann sollten Sie das lieber in zwei Fragen aufteilen und das „mehr“ weglassen. Hier soll nur veranschaulicht werden, dass sich eine Frage auf unterschiedliche Weise verbessern lässt und dass von der Frage „Was möchte ich messen?“, geleitet sein sollte.

Darüber hinaus sollten wir zwar versuchen, uns kurz zu fassen, dabei aber präzise sein. Ein positives Beispiel (ebenfalls aus Reitstätter et al. 2020): „Mit wie vielen Personen (außer Ihnen selbst) haben Sie heute das Museum besucht?“ Hier ist die Einbeziehung von „außer Ihnen selbst“ entscheidend, da die Frage sonst unklar wäre. Dies würde zu Problemen für die Forschenden führen, da sie am Ende nicht wüssten, ob ‚zwei‘ bedeutet, dass der:die Teilnehmende die Ausstellung mit ein oder zwei anderen Personen besucht hat.

Schritt 4: Letztlich müssen Sie sich für Antwortmöglichkeiten entscheiden. Allgemein gibt es drei Arten von Antwortoptionen: dichotom (z. B. ja/nein, wahr/falsch), kategorial (z. B. Nationalität) oder kontinuierlich (z. B. eine 7-Punkte-Skala). Kategoriale Antworten können schwer zu analysieren sein, wenn sie als abhängige Variable verwendet werden, sie werden daher meist für Fragen genutzt, die sich auf beschreibende Aspekte der Stichprobe beziehen (z. B. Nationalität). Wenn Sie mehr als zwei Antwortoptionen verwenden, ist es wichtig, dass die Antwortkategorien gleichmäßig verteilt sind. Geschieht dies nicht, zum Beispiel bei einer 4-Punkte-Skala, bei der 1 ‚stimme nicht zu‘, 2 ‚stimme zu‘, 3 ‚stimme stark zu‘ und 4 ‚stimme extrem zu‘ bedeutet, können Sie Zustimmung gut messen (mit hoher Messgenauigkeit), aber Ablehnung nicht ausreichend erfassen. Natürlich ist das Ziel, eine möglichst umfassende Messung zu erreichen, sodass dies suboptimal ist (Wenig, 2004). Als Faustregel werden mindestens 5- oder 7-Punkte-Skalen empfohlen, um die Daten als kontinuierlich behandeln zu können (Wenig, 2004). Das bedeutet jedoch nicht, dass eine höhere Anzahl von Skalenpunkten immer besser ist. Tatsächlich kann dies die Validität verringern, da die Teilnehmenden möglicherweise nicht in der Lage sind, die erforderlichen subtilen Unterscheidungen zu treffen (Clark und Watson 1995, Symonds 1924, Wenig 2004). Schließlich wird manchmal eine gerade Anzahl von Skalenpunkten bevorzugt, um die potenziell schwierige Interpretation des mittleren Skalenwerts zu vermeiden. Dies könnte jedoch dazu führen, dass die Befragten Antworten geben müssen, die nicht ihrer tatsächlichen Meinung oder ihren Gefühlen entsprechen, was auch problematisch sein kann (Clark und Watson 1995).

Es gibt verschiedene Möglichkeiten, Fragebögen anzubieten. Sie können Papier und Stift verwenden (z. B. Pelowski et al. 2022), Tablets nutzen (z. B. Reitstätter et al. 2020) oder die Teilnehmenden können ihr eigenes Handy verwenden (z. B. Specker et al. 2020). Im letzteren Fall können die Teilnehmenden einen QR-Code scannen, der den Fragebogen oder Link auf ihrem Telefon öffnet. Ein Nachteil von Papier und Stift ist, dass die Antworten manuell in z. B. eine Excel-Datei eingegeben werden müssen, um die Daten zu analysieren. Dies kann viel Zeit kosten und durch einfache menschliche Versehen zu Fehlern in den Daten führen. Dennoch handelt es sich hier um ein einfaches Verfahren, das weder eine technische Infrastruktur (z. B. Lademöglichkeit für Geräte, Softwarelizenzen) noch Programmierkenntnisse erfordert.

Was den personellen Aufwand betrifft, können Forschende im Prinzip alle Schritte im Alleingang durchführen: Die Studie entwerfen, die Fragen auswählen oder erstellen, die Daten erheben und analysieren. Erforderlich sind allerdings Fachkenntnisse im experimentellen Design für quantitative Studien sowie in der statistischen Datenanalyse. Je nach Sachkenntnis und Erfahrung des/der Forschenden mit quantitativen Methoden sowie der relativen Komplexität der Untersuchung variiert der Zeitaufwand für die einzelnen Schritte. Außerdem müssen Schritt 2 bis 4 nicht durchgeführt werden, wenn in Schritt 1 eine passende, validierte Skala gefunden werden kann, was Zeit sparen würde.

Die Dauer der Datenerhebung hängt wahrscheinlich nicht von den Forschenden ab, sondern von anderen Faktoren, wie der angestrebten Stichprobengröße (d. h., wie viele Personen getestet werden sollen) und der Anzahl der Personen, die teilnehmen können. Beispielsweise könnten Sie in einem großen Museum wie der Albertina in Wien innerhalb weniger Tage 100 Personen testen (z. B. Specker et al. 2020), während Sie in kleineren Museen mit weniger Besuchenden pro Tag einen längeren Zeitraum benötigen würden. Abhängig davon, wie lang der Fragebogen ist, kann die Teilnahme relativ wenig Zeit beanspruchen. Hierbei ist zu berücksichtigen, ob Sie reguläre Museumsbesucher:innen testen, oder ob Sie Personen für ihre Untersuchung gezielt ins Museum bringen. Im ersten Fall sollten Sie bedenken, dass die Teilnehmenden das Museum besuchen, um es zu besichtigen, und nicht, um an Ihrer Studie teilzunehmen. Daher gilt: Je kürzer Sie den Fragebogen halten können, desto besser. Im zweiten Fall können Sie von Ihren Teilnehmenden in der Regel mehr verlangen, da sie speziell zur Teilnahme an Ihrer Studie ins Museum gekommen sind.

Anwendungsbeispiel

Ein Beispiel für eine Studie, die Fragebögen in einem Ausstellungskontext verwendete, ist Specker et al. (2020). In dieser Studie untersuchten wir das kuratorische

Narrativ – also die Einbettung von Kunstwerken oder einer gesamten Ausstellung in einen größeren Kontext von Bedeutung und Relevanz. In der Studie besuchten die Hälfte der Teilnehmenden die Monet-Retrospektive im Albertina Museum in Wien, während die andere Hälfte die Dauerausstellung *Monet bis Picasso* besuchte. In beiden Ausstellungen wurden die Teilnehmenden gebeten, den *Seerosenteich* (1917–1919) von Claude Monet zu betrachten. In der Retrospektive hing dieses Gemälde in einem Raum, der den stilistischen Wandel in Monets Werk hin zu abstrakteren Malereien dokumentierte. In der Dauerausstellung war das Werk in einem Raum mit anderen impressionistischen Kunstwerken ausgestellt. Während in der Retrospektive eine stilistische Abweichung vorlag, sich also das Kunstwerk von den anderen Werken im Raum hinsichtlich seines Stils unterschied, war dies in der Dauerausstellung nicht der Fall. Frühere Studien (z. B. Stamkou, van Kleef und Homan 2018) hatten bereits gezeigt, dass Künstler:innen von den Betrachtenden als einflussreicher eingeschätzt werden, wenn deren Werk in einem Kontext stilistischer Abweichung präsentiert wird. Das konnten wir in unserer Ausstellungsstudie ebenfalls bestätigen: Monet wurde in der Retrospektive als einflussreicher wahrgenommen als in der Dauerausstellung. In diesem Fall wurden sowohl die wahrgenommene Bedeutung als auch die Abweichung durch einen Fragebogen gemessen. Da wir unsere Studie auf Stamkou et al. (2018) stützten, konnten wir deren Fragen übernehmen und lediglich an unseren Ausstellungskontext anpassen. Zum Beispiel fragten wir anstelle von „Was denken Sie über diesen Künstler bzw. diese Künstlerin?“ spezifisch: „Was denken Sie über Monet als Maler?“ Anschließend folgten dieselben vier Aussagen wie bei Stamkou et al. (2018), jedoch mit der Anpassung, dass „Künstler“ durch „Monet“ ersetzt wurde, z. B.: „Ich denke, dass [Monet] auch nach vielen Generationen von Malern einen großen Beitrag zur Kunst leisten wird.“

Diese Studie führte zu mehreren Erkenntnissen: Erstens stützt sie eine lange gehegte Annahme in der kuratorischen Praxis – nämlich, dass die Entscheidung der Kurator:innen bei der Erstellung eines Gesamtnarrativs für eine Ausstellung (insbesondere die Reihenfolge, in der Werke betrachtet werden) tatsächlich zu messbaren Unterschieden führt. Zweitens können diese kuratorischen Narrative unsere Sichtweise auf Künstler:innen prägen – selbst bei sehr bekannten und berühmten Künstlern wie Monet. Möglicherweise wirkt sich das bei weniger bekannten Künstler:innen sogar noch stärker aus. Insgesamt könnten Studien wie diese die Grundlage für einen evidenzbasierten Ansatz für die Museums-Kuratierung bilden, bei dem Kurator:innen empirische Erkenntnisse nutzen, um kuratorische Ziele zu erreichen oder anzupassen und das Ausstellungserlebnis der Besucher:innen zu gestalten.

Methodenreflexion

Wie in der Einleitung erwähnt und oben diskutiert, haben Fragebögen den Vorteil, dass sich mit ihnen das subjektive Ausstellungserlebnis direkt erfassen lässt. Sie sind kostengünstig und zeiteffizient, flexibel in der Art, welche Fragen wie gestellt werden können, und relativ einfach umzusetzen (Simms 2008). Der Fokus auf quantitative Messungen hat den Vorteil, dass Fragebögen statistische Analysen und relativ einfache Interpretationen ermöglichen – ein Durchschnittswert ist eine aussagekräftige Zusammenfassung der Antworten verschiedener Befragter. Es wäre weitaus schwieriger, eine derart kurze Zusammenfassung aus phänomenologischen Interviews oder anderen qualitativen Antworten zu erstellen.

Allerdings basieren diese Methoden (sowohl Fragebögen als auch qualitative Methoden) auf Selbstbeobachtung. Ihnen liegt die Annahme zugrunde, dass Teilnehmende Einblick in ihre subjektiven Erfahrungen haben und diese durch ihre Antworten auf die gestellten Fragen wiedergeben können. Es ist ein Vorteil der quantitativ orientierten Fragebögen, dass es Teilnehmenden im Allgemeinen leichter fällt, auf die Frage ‚Wie schön finden Sie dieses Kunstwerk?‘ auf einer festgelegten Skala zu antworten, z. B. von 1 (überhaupt nicht schön) bis 7 (sehr schön), anstatt auf eine offene Frage, bei der die Teilnehmenden ihre Gedanken verbal ausdrücken müssten. Das könnte je nach Thema und Zielgruppe schwierig sein. Ein Nachteil ist jedoch, dass die Befragten in ihren Antworten nicht flexibel sind – man erhält nur Antworten auf die Fragen, die gestellt wurden, was zwangsläufig zu einem begrenzten Fokus führt. Wenn die Fragen kurz sind und die Antwortmöglichkeiten eingeschränkt, erhalten wir zudem keine so reichen Datensätze wie bei qualitativen Methoden. Zum Beispiel erfahren wir vielleicht, ob etwas gemocht wird, aber nicht unbedingt warum, zumindest nicht auf individueller Ebene.

Darüber hinaus können Fragebögen (oder andere Methoden, bei denen Teilnehmende direkt befragt werden) an Validität verlieren, wenn verschiedene Arten von Antwortverzerrungen auftreten. Befragte könnten sich beispielsweise veranlasst sehen, nicht ganz ehrlich zu sein, wenn es etwa um sozial heikle Themen geht. Jemanden zu fragen ‚Sind Sie rassistisch?‘ aber vielleicht auch ‚Interessieren Sie sich für Kunst?‘ (z. B. wenn dies in einem kunstbezogenen Kontext wie einem Museum gefragt wird), ist möglicherweise nicht so empfehlenswert. Allerdings fanden McKibben und Silvia (2017) keinen Hinweis auf solche Verzerrungen von sozialer Erwünschtheit in Antworten auf Kreativitäts- und Kunstskaleten.

Abschließend, nach der Diskussion all dieser Aspekte von Fragebögen in der Kunstforschung, möchten wir betonen, dass der Nutzen von quantitativen Fragebögen am größten ist, wenn Forschende klare Vorstellungen von den Hypothesen haben, die sie testen möchten. Trotz der Versuchung, verschiedene Komponenten hinzuzufügen, ist es immer wichtig zu entscheiden, welche statistischen Analysen durchgeführt werden sollen, um diese in konfirmatorischen Studien zu testen. Mit

diesen Anforderungen im Blick sollten Sie für die Ausarbeitung guter, geeigneter Fragen für eine sinnvolle, wissenschaftliche Ausstellungsanalyse gut gewappnet sein!

Literaturverzeichnis

- Benson, Jeri und Dennis Hocevar. 1985. The Impact of Item Phrasing on the Validity of Attitude Scales for Elementary School Children. In *Journal of Educational Measurement* 22 (3): 231–40. <https://sci-hub.st/10.2307/1435036> (05.08.2024).
- Clark, Lee Anna und David Watson. 1995. Constructing Validity: Basic Issues in Objective Scale Development. *Psychological Assessment* 7 (3): 309–319. <http://www.bwgriffin.com/gsu/courses/edur9131/2018spr-content/04-questionnaire/04-Clark-1995.pdf> (05.08.2024).
- Gehlbach, Hunter und Maureen E. Brinkworth. 2011. Measure Twice, Cut down Error: A Process for Enhancing the Validity of Survey Scales. *Review of General Psychology* 15 (4): 380–387. <https://doi.org/10.1037/a0025704>.
- Gerger, Gernot, Helmut Leder, Pablo P. L. Tinio und Annekathrin Schacht. 2011. Faces versus Patterns: Exploring Aesthetic Reactions Using Facial EMG. In *Psychology of Aesthetics, Creativity, and the Arts* 5 (3): 241–250. <https://doi.org/10.1037/a0024154>.
- Kuchinke, Lars, Sabrina Trapp, Arthur M. Jacobs und Helmut Leder. 2009. Pupillary Responses in Art Appreciation: Effects of Aesthetic Emotions. In *Psychology of Aesthetics, Creativity, and the Arts* 3 (3): 156–163. <https://doi.org/10.1037/A0014464>.
- Leder, Helmut, Aleksandra Mitrovic und Jürgen Goller. 2016. How Beauty Determines Gaze! Facial Attractiveness and Gaze Duration in Images of Real World Scenes. In *I-Perception* 7 (4). <https://doi.org/10.1177/2041669516664355>.
- McKibben, William Bradley und Paul J. Silvia. 2017. Evaluating the Distorting Effects of Inattentive Responding and Social Desirability on Self-Report Scales in Creativity and the Arts. *The Journal of Creative Behavior* 51 (1): 57–69. <https://doi.org/10.1002/jocb.86>.
- Pelowski, M., Eva Specker, Jane Boddy, Beatrice Immelmann, Felix Haiduk, Giovanni Spezie, Paula Ibáñez de Aldecoa, Hillary Jean-Joseph, Helmut Leder und Patrick S. Markey. 2022. Together in the Dark? Investigating the Understanding and Feeling of Intended Emotions Between Viewers and Professional Artists at the Venice Biennale. In *Psychology of Aesthetics, Creativity, and the Arts*. <https://doi.org/10.1037/ACA0000436>.
- Reitstätter, Luise, Hanna Brinkmann, Thiago Santini, Eva Specker, Zoya Dare, Flora Bakondi, Anna Miscená, Enkelejda Kasneci, Helmut Leder und Raphael Rosenberg. 2020. The Display Makes a Difference: A Mobile Eye Tracking Study on the

- Perception of Art before and after a Museum's Rearrangement. In *Journal of Eye Movement Research* 13 (2). <https://doi.org/10.16910/jemr.13.2.6>.
- Schindler, Ines, Georg Hosoya, Winfried Menninghaus, Ursula Beermann, Valentin Wagner, Michael Eid und Klaus R. Scherer. 2017. Measuring Aesthetic Emotions: A Review of the Literature and a New Assessment Tool. In *PLOS ONE* 12 (6): e0178899. <https://doi.org/10.1371/JOURNAL.PONE.0178899>.
- Simms, Leonard J. 2008. Classical and Modern Methods of Psychological Scale Construction. In *Social and Personality Psychology Compass* 2 (1): 414–433. <https://doi.org/10.1111/j.1751-9004.2007.00044.x>.
- Specker, Eva. 2021. Further Validating the VAIK: Defining a Psychometric Model, Configural Measurement Invariance, Reliability, and Practical Guidelines. In *Psychology of Aesthetics, Creativity, and the Arts*. <https://doi.org/10.1037/ACA0000427>.
- Specker, Eva, Katherine N. Cotter und Kyung Yong Kim. 2023. The next Step for the VAIK: An Item-Focused Analysis. In *Psychology of Aesthetics, Creativity, and the Arts*. <https://doi.org/10.1037/ACA0000559>.
- Specker, Eva, Michael Forster, Hanna Brinkmann, Jane Boddy, Matthew Pelowski, Raphael Rosenberg und Helmut Leder. 2020. The Vienna Art Interest and Art Knowledge Questionnaire (VAIK): A Unified and Validated Measure of Art Interest and Art Knowledge. *Psychology of Aesthetics, Creativity, and the Arts* 14 (2): 172–185. <https://doi.org/10.1037/aca0000205>.
- Specker, Eva, Eftychia Stamkou, Matthew Pelowski und Helmut Leder. 2020. Radically Revolutionary or Pretty Flowers? An Experimental Museum Study of the Impact of Curatorial Narrative Highlighting Artistic Deviance on the Visitor's Assessment of Artist Influence. In *Psychology of Aesthetics Creativity and the Arts*. <https://doi.org/https://doi.org/10.1037/aca0000320>.
- Stamkou, Eftychia, Gerben A. van Kleef und Astrid C. Homan. 2018. The Art of Influence: When and Why Deviant Artists Gain Impact. In *Journal of Personality and Social Psychology* 115 (2): 276–303. <https://doi.org/10.1037/pspi0000131>.
- Swain, Scott D., Danny Weathers und Ronald W. Niedrich. 2008. Assessing Three Sources of Misresponse to Reversed Likert Items. In *Journal of Marketing Research* 45 (1): 116–131. <https://doi.org/10.1509/JMKR.45.1.116>.