

## Eine Kurzsкала zur Erfassung des Gamma-Faktors sozial erwünschten Antwortverhaltens:

Die Kurzsкала Soziale Erwünschtheit-Gamma  
(KSE-G)

*Christoph J. Kemper, Constanze Beierlein,  
Doreen Bensch, Anastassiya Kovaleva &  
Beatrice Rammstedt*



GESIS-Working Papers 2012|25

## **Eine Kurzsкала zur Erfassung des Gamma-Faktors sozial erwünschten Antwortverhaltens:**

Die Kurzsкала Soziale Erwünschtheit-Gamma (KSE-G)

*Christoph J. Kemper, Constanze Beierlein, Doreen Bensch,  
Anastassiya Kovaleva & Beatrice Rammstedt*

## GESIS-Working Papers

GESIS – Leibniz-Institut für Sozialwissenschaften

Postfach 12 21 55

68072 Mannheim

Telefon: (0621) 1246 - 253

Telefax: (0621) 1246 - 500

E-Mail: [constanze.beierlein@gesis.org](mailto:constanze.beierlein@gesis.org)

ISSN: 1869-0491 (Online)

Herausgeber,

Druck und Vertrieb:

GESIS - Leibniz-Institut für Sozialwissenschaften  
Unter Sachsenhausen 6-8, 50667 Köln

# 1 Einleitung

---

Erhebungsinstrumente zur Erfassung von psychologischen Merkmalen, wie beispielsweise Persönlichkeit, Risikobereitschaft, Werte, Lebenszufriedenheit, Attraktivität, Optimismus oder Intelligenz, werden immer häufiger in der Forschung eingesetzt. Neben der psychologischen Forschung, in der psychologische Merkmale traditionell im Zentrum des Interesses stehen, werden psychologische Merkmale vermehrt auch in anderen Forschungsbereichen erhoben, zum Beispiel in den Sozialwissenschaften, den Wirtschaftswissenschaften, den Ingenieurwissenschaften und den Gesundheitswissenschaften. Die Forscherinnen und Forscher der verschiedenen Disziplinen erwarten von der Erfassung dieser Merkmale in Umfragen, Studien und Experimenten, und deren Implementierung in ihre Modelle, eine bessere Deskription und Prädiktion wissenschaftlich und gesellschaftlich relevanter Prozesse und Phänomene.

Der Zusammenhang zwischen psychologischen Merkmalen und sozioökonomischen Erfolgsgrößen, die bei Sozialwissenschaftlern und Ökonomen im Fokus des Interesses stehen, gilt als gut gesichert (Gottfredson, 1997; Gottfredson & Deary, 2004; Schmidt & Hunter, 1998; Strenze, 2007). Diese Studien belegen, dass die kognitiven Fähigkeiten einer Person, insbesondere die Intelligenz, der beste Prädiktor für ein erfolgreiches Leben ist. Personen mit hoher kognitiver Leistungsfähigkeit haben im Vergleich zu Personen mit niedriger kognitiver Leistungsfähigkeit mehr Erfolg in Schule, Studium, Ausbildung, Beruf und im Privatleben. Personen mit hoher kognitiver Leistungsfähigkeit haben zum Beispiel meist ein höheres Einkommen oder eine höhere Position im Beruf, lassen sich seltener scheiden und werden seltener delinquent oder arbeitslos. Neben der kognitiven Leistungsfähigkeit sind auch andere psychologische Merkmale dem sozioökonomischen Erfolg zuträglich, wie zum Beispiel Gewissenhaftigkeit und Optimismus (Barrick & Mount, 1991; Kemper, Kovaleva, Beierlein & Rammstedt, im Druck).

Neben diversen sozioökonomischen Erfolgsgrößen beeinflussen psychologische Merkmale viele Prozesse und Phänomene, die mitunter weitreichende Implikationen für den Einzelnen, seine Mitmenschen oder die Gesellschaft als Ganzes haben. So kann beispielsweise auf Grundlage von psychologischen Merkmalen das Wahlverhalten prädiziert werden (Schumann & Schoen, 2005). Auch auf das Gesundheitsverhalten und dessen Folgen haben psychologische Merkmale einen Einfluss. Befunde aus der Fachliteratur zeigen, dass Merkmale wie Gewissenhaftigkeit und Optimismus die physische und die psychische Gesundheit, inklusive der Morbidität und der Mortalität beeinflussen können (Allison, Guichard, Fung & Gilain, 2003; Arthur & Graziano, 1996; Rasmussen, Scheier & Greenhouse, 2009). Schließlich stehen psychologische Merkmale im Zusammenhang mit der Entwicklung und Aufrechterhaltung psychischer Störungen, zum Beispiel Depression und Phobien, und mit Drogen- und Alkoholkonsum sowie delinquentem Verhalten (Block, Block & Keyes, 1988; Block, Gjerde & Block, 1991; Gottfredson, 1997).

Aufgrund dieser vielfältigen Beziehungen und der Nützlichkeit von psychologischen Merkmalen zur Verbesserung der Deskription und Prädiktion wissenschaftlich und gesellschaftlich relevanter Prozesse und Phänomene, forderte kürzlich auch der Ökonomie-Nobelpreisträger James Heckman, dass zukünftige sozialwissenschaftliche Studien vermehrt validierte Persönlichkeitsskalen und Intelligenztests umfassen sollten (Borghans, Duckworth, Heckman & ter Weel, 2008). Diesen Standpunkt wird auch von anderen Forscherinnen und Forschern (Goldberg, 2005; Rammstedt, 2010a) und Institutionen (Rat für Sozial- und Wirtschaftsdaten, 2010) vertreten. Seit einigen Jahren sind diesbezüglich deutliche Bestrebungen erkennbar. Im Sozio-ökonomischen Panel (SOEP) wurden in 2004 erstmals Risikoaversion und in 2005 Persönlichkeit, Kontrollüberzeugung und Reziprozität erfasst; in 2006 auch grundlegende Intelligenzmaße. Darüber hinaus wurden Skalen zur Erfassung psychologischer Merkmale in weitere Untersuchungen aufgenommen, zum Beispiel International Social Survey Programme (ISSP), Household, Income and Labour Dynamics in Australia (HILDA), UK Household Longitudinal Study (UKHLS)

und DNB Household Survey (DHS). Der Bedarf an Verfahren zur Operationalisierung psychologischer Merkmale ist demnach gegeben und wird in den kommenden Jahren vermutlich weiter steigen.

Forscherinnen und Forscher, die entsprechende Merkmale in ihrer Untersuchung erfassen möchten, stehen allerdings zurzeit meist vor dem Problem, für ihre Zwecke geeignete Erhebungsinstrumente zu finden. Aus der psychologischen Forschung sind viele Erhebungsinstrumente bekannt (Leibniz-Zentrum für Psychologische Information und Dokumentation, 2011). Diese sind auch in den meisten Fällen sorgfältig im Hinblick auf ihre psychometrische Güte geprüft. Das heißt, durch empirische Untersuchungen wurde festgestellt, dass die Erhebungsinstrumente ein bestimmtes psychologisches Merkmal (Konstrukt) präzise messen (Gütekriterium der Reliabilität oder Messgenauigkeit), und dass sie das Merkmal, das sie zu messen intendieren, auch tatsächlich messen (Gütekriterium der Gültigkeit oder Validität). Allerdings beinhalten diese Erhebungsinstrumente oft viele Fragen oder Aufgaben (Items), um die Konstrukte zu erfassen. Während in der psychologischen Forschung, insbesondere aber in der psychologischen Einzelfalldiagnostik, eine möglichst breite Erfassung der jeweiligen Merkmale notwendig ist und die Dauer der Erhebung (Gütekriterium Ökonomie) eine untergeordnete Rolle spielt, ist die Erhebungsdauer in den meisten sozialwissenschaftlichen Untersuchungen ein kritischer Kostenfaktor. Hier spielt die Anzahl der Items bzw. die Länge von Erhebungsinstrumenten sehr wohl eine Rolle. Da die Erhebungsinstrumente für psychologische Merkmale zwar nicht im Zentrum des Interesses stehen, aber aufgrund ihrer Nützlichkeit dennoch mit erhoben werden sollen, werden oft auf der Grundlage etablierter psychologischer Skalen und Tests gekürzte Ad-Hoc-Instrumente erstellt. Der Einsatz solcher Ad-Hoc-Instrumente kann in manchen Fällen einen Mehrwert für eine bestimmte Untersuchung darstellen, bringt allerdings den Nachteil mit sich, dass Befunde zwischen Untersuchungen nicht mehr vergleichbar sind, wenn die eingesetzten Erhebungsinstrumente durch unterschiedliche Kürzungsstrategien zustande gekommen sind. Durch eine Kürzung sind genau genommen auch die Belege der psychometrischen Güte, die für das originäre Erhebungsinstrument empirisch ermittelt wurden, nicht mehr gültig. Die psychometrische Güte müsste erneut durch ein Wechselspiel von konzeptuellen Überlegungen und empirischen Prüfungen (Validierung) belegt werden (Stanton, Sinar, Balzer & Smith, 2002). Der Aufwand für den einzelnen Forscher wäre immens.

Die Bereitstellung von Beratungs- und Serviceleistungen für die Sozialwissenschaften gehört zu den Kernaufgaben von GESIS. Dies beinhaltet auch die Entwicklung und Verbreitung von geprüften, qualitativ hochwertigen und standardisierten Erhebungsinstrumenten. Bisher liegen für die Erfassung von für die sozialwissenschaftliche Forschung relevanten psychologischen Merkmalen noch keine etablierten und disziplinübergreifend akzeptierten Erhebungsinstrumente vor, die eine ökonomische und effiziente Messung erlauben. Ziel der hier beschriebenen Skalenentwicklung war es daher, für das Konstrukt Optimismus-Pessimismus ein angemessenes Erhebungsinstrument zu entwickeln, es umfassend zu validieren und anschließend potentiellen Nutzern zur Verfügung zu stellen ([www.gesis.org/kurzskalen-psychologischer-merkmale](http://www.gesis.org/kurzskalen-psychologischer-merkmale)). Damit ist die Hoffnung verknüpft, dass durch den vermehrten Einsatz dieses standardisierten psychologischen Erhebungsinstruments eine erhöhte Anschlussfähigkeit und Vergleichbarkeit zwischen Untersuchungen und eine verbesserte Deskription und Prädiktion wissenschaftlich und gesellschaftlich relevanter Prozesse und Phänomene erzielt werden kann.

## 2 Skalenkonzept

---

### 2.1 Theoretischer Hintergrund

Seit mehr als 50 Jahren wird das Konstrukt Soziale Erwünschtheit in der psychologischen Forschung untersucht (Crowne & Marlowe, 1960; Edwards, 1957; Paulhus, 1991, 2002). Am weitesten verbreitet ist die Konstruktdefinition von Paulhus (2002) nach der soziale Erwünschtheit als Tendenz verstanden wird überwiegend positive Beschreibungen der eigenen Person abzugeben („tendency to give overly positive self-descriptions“, S. 50). Soziale Erwünschtheit hat weitreichende Implikationen für die Messung von Persönlichkeitsmerkmalen oder Einstellungen über Fragebögen oder Interviews (Selbstberichte). Auf der Verhaltensebene schlägt sich Soziale Erwünschtheit in Antwortverzerrungen (response bias) nieder, die beispielsweise zu einer Konfundierung von statistischen Zusammenhängen zwischen Untersuchungsvariablen führen können. In den vergangenen Dekaden wurden zahlreiche Studien durchgeführt, um die Struktur des Konstrukts zu untersuchen, geeignete Messinstrumente zu konstruieren und diese mit dem Ziel der statistischen Kontrolle der Konfundierung einzusetzen (für eine Übersicht solcher Arbeiten siehe zum Beispiel Paulhus, 1991, 2002).

Verschiedene Modelle zur Strukturierung des Konstrukts wurden während der letzten Jahre vorgeschlagen (für eine Übersicht siehe Paulhus, 2002). Paulhus präsentiert das bislang umfassendste Modell, das als Synthese aus bisherigen empirischen Befunden und konzeptuellen Überlegungen angesehen werden kann. Er schlägt ein mehrdimensionales Modell des Konstrukts Soziale Erwünschtheit vor. Auf der obersten Ebene der hierarchischen Struktur sind zwei abstrakte Faktoren angeordnet: Alpha und Gamma. Gemeinsam ist diesen Faktoren, dass sie sowohl einen Selbsttäuschungsstil als auch einen Fremdtäuschungsstil beinhalten. Allerdings sollen sich die Faktoren Alpha und Gamma im Hinblick auf ihre Inhalte unterscheiden. Nach Paulhus gehen Unterschiede in den beiden Faktoren auf zwei zugrunde liegende Werte zurück: Agency (Selbstbehauptung, Selbstschutz, Selbstentfaltung, Abgrenzung von anderen, Streben nach Macht und Kontrolle) und Communion (Gemeinsinn, Verbundenheit, Offenheit für die Gefühle anderer, Streben nach Gemeinschaft, Teilnahme, Kooperation und Bindung; vgl. Bakan, 1966). Wenn Personen Agency-Werte stark internalisiert haben, dann zeigen sie laut Paulhus einen egoistischen Bias (mit Alpha assoziiert). Diese Personen neigen dazu ihren sozialen und intellektuellen Status zu übertreiben. Sie geben unrealistisch positive Selbstbeschreibungen zu Persönlichkeitsmerkmalen ab, die mit Agency assoziiert sind, zum Beispiel zu Dominanz, Angstfreiheit, Intellekt und Kreativität. Neben diesen weitgehend unbewussten Prozessen (Selbsttäuschung), die sich in einem egoistischen Bias manifestieren können, nennt Paulhus auch bewusste Versuche der Eindruckssteuerung (Fremdtäuschung), zum Beispiel prahlen und angeben, die er als Agency-Management bezeichnet. Bei Personen, die Communion-Werte stark internalisiert haben, wird angenommen, dass sie einen moralistischen Bias zeigen (mit Gamma assoziiert). Diese Personen neigen dazu sozial unerwünschte Impulse zu leugnen und sich positive Attribute zuzuschreiben. Sie geben unrealistisch positive Selbstbeschreibungen zu Persönlichkeitsmerkmalen ab, die mit Communion assoziiert sind, zum Beispiel zu Verträglichkeit, Pflichtbewusstsein und Selbstbeherrschung. Auf Ebene der Fremdtäuschung sollen sie dazu neigen in Situationen, die ihre Communion-Werte ansprechen, schadensbegrenzendes Verhalten zu zeigen und Ausreden vorzubringen, um in einem besseren Licht zu erscheinen. Dieses von Paulhus vorgestellte Modell führt bisherige Ansätze überzeugend in einem integrativen Modell zusammen (für weitere Details siehe Paulhus, 2002). Die bisherigen Dichotomien wie Selbsttäuschung vs. Fremdtäuschung oder bewusstes vs. unbewusstes Verhalten wurden aufgelöst bzw. in die abstrakten Faktoren Alpha und Gamma integriert: Beide Faktoren sind mit einem Bias assoziiert, der sich aus internalisierten Werten speist und sowohl aufgrund von bewussten als auch unbewussten Prozessen in sozial erwünschtem (Antwort-)Verhalten manifestieren kann.

Für die Umfrageforschung ist insbesondere eine Messung des Gamma-Faktors der sozialen Erwünschtheit von Interesse. In der Umfrageforschung werden Daten primär per Face-to-face-Interview erhoben. Diese besondere Interaktionssituation zwischen Interviewer und Befragungsperson regt Communion-Werte auf Seiten der Befragungsperson an. Weiterhin verweisen Interviewer in sozialwissenschaftlichen Umfragen, in Abgrenzung zu kommerziell motivierten Marktforschungsumfragen, nicht selten auf die gesellschaftliche Bedeutung der Umfrage, um die Teilnahmebereitschaft zu erhöhen. Bei der Befragungsperson könnte dies einen durch die Situation ohnehin induzierten moralistischen Bias weiter verstärken. Er könnte dazu führen, dass sich die Befragungsperson bemüht wie eine „nette Person“ oder ein „guter Bürger“ auf die Fragen zu antworten. Ihr Antwortverhalten wird dadurch verzerrt. Um diesen möglichen Einflussfaktor auf empirische Untersuchungsergebnisse ökonomisch quantifizieren zu können, wurde eine Kurzsкала zur Erfassung des Gamma-Faktors der Sozialen Erwünschtheit (KSE-G) entwickelt und validiert.

## 2.2 Aufbau

Die KSE-G erfasst zwei Aspekte des Gamma-Faktors von Soziale Erwünschtheit, die Übertreibung positiver Qualitäten (PQ+) und die Minimierung negativer Qualitäten (NQ-) mit jeweils drei Items (siehe Abschnitt 2.4 und Appendix B). Für die Antworten der Befragungsperson steht eine fünfstufige Ratingskala zur Verfügung. Die Antwortkategorien der KSE-G gehen von „trifft gar nicht zu“ (0) bis „trifft voll und ganz zu“ (4).

## 2.3 Auswertung

Um einen Messwert (Skalenwert) für die individuelle Ausprägung der Befragungsperson in PQ+ bzw. NQ- zu erhalten, werden deren Antworten auf den drei Items der jeweiligen Skala aggregiert: Die drei Items werden zunächst summiert und anschließend durch die Anzahl der Items pro Skala geteilt. Der Wertebereich für PQ+ und NQ- liegt zwischen 0 und 4 (für Referenzwerte siehe Abschnitt 5.4 und Appendix A).

## 2.4 Items

Die sechs Items der KSE-G und deren deskriptive Statistiken sind in Tabelle 1 dargestellt. Letztere basieren auf Daten einer umfangreichen, bevölkerungsrepräsentativen Zufallsstichprobe (Stichprobe 2, siehe Tabelle 2). Um die Nutzung der Skala für englischsprachige Untersuchungen zu ermöglichen, wurde diese ins Englische übersetzt. Die International Test Commission (2010) empfiehlt hierbei ein zweistufiges Verfahren. Zunächst übersetzten zwei hauptberufliche, muttersprachliche Übersetzer die Items unabhängig voneinander. Dabei wurde eine Übersetzung in britischem Englisch und die andere in amerikanischem Englisch angefertigt. In der zweiten Phase des Übersetzungsprozesses fand ein Rekonziliationstreffen statt, in dessen Verlauf die Übersetzungsvorschläge in einer Gruppe von Experten für die psychologischen Merkmale, den Übersetzern und einem weiteren Experten für Fragebogenübersetzung diskutiert und überarbeitet wurden. Die englische Version der Skala ist in Appendix C zu finden. Die Güte der Übersetzungen wurde bislang allerdings noch nicht geprüft.

Tabelle 1: Items und deskriptive Statistiken der KSE-G aus Stichprobe 2.

	<i>M</i>	<i>SD</i>	Sch	Kurt
Im Streit bleibe ich stets sachlich und objektiv.	2.55	1.01	-0.41	-0.34
Auch wenn ich selbst gestresst bin, behandle ich andere immer freundlich und zuvorkommend.	2.72	0.99	-0.57	-0.13
Wenn ich mich mit jemandem unterhalte, höre ich ihm immer aufmerksam zu.	3.04	0.87	-0.98	1.30
<i>Subskala Übertreibung positiver Qualitäten (PQ+)</i>	2.66	0.86	-0.43	-0.16
Es ist schon mal vorgekommen, dass ich jemanden ausgenutzt habe.	0.93	1.01	0.96	0.28
Ich habe schon mal Müll einfach in die Landschaft oder auf die Straße geworfen.	0.59	1.00	1.77	2.39
Manchmal helfe ich jemandem nur, wenn ich eine Gegenleistung erwarten kann.	0.76	0.98	1.21	0.75
<i>Subskala Untertreibung negativer Qualitäten (NQ-)</i>	3.24	0.78	-1.04	0.58

Anmerkungen: Sch = Schiefe, Kurt = Kurtosis. *N* = 566.

Um die Nutzung der Skala für englischsprachige Untersuchungen zu ermöglichen, wurde diese ins Englische übersetzt. Die International Test Commission (2010) empfiehlt hierbei ein zweistufiges Verfahren. Zunächst haben zwei hauptberufliche, muttersprachliche Übersetzer die Items unabhängig voneinander übersetzt. Dabei wurden eine Übersetzung in britischem Englisch und die andere in amerikanischem Englisch angefertigt. In der zweiten Phase des Übersetzungsprozesses fand ein Rekonziliationstreffen statt, in dessen Verlauf die Übersetzungsvorschläge in einer Gruppe von Experten für die psychologischen Merkmale, den Übersetzern und einem weiteren Experten für Fragebogenübersetzung diskutiert und überarbeitet wurden. Die englische Version der Skala ist in Appendix C zu finden. Die Güte der Übersetzungen wurde bislang allerdings noch nicht geprüft.

### 3 Durchführung

---

Die KSE-G wurde als Forschungsinstrument für sozialwissenschaftliche Untersuchungen unterschiedlichster Art und Fragestellung entwickelt. Als Zielgruppe wurde daher die deutschsprachige Allgemeinbevölkerung ab 18 Jahren gewählt<sup>1</sup>. Die Itemformulierungen wurden in einem Expertenreview für die Zielgruppe optimiert. Auch die in Kapitel 5 berichteten empirisch ermittelten Gütekriterien beziehen sich auf diese Zielgruppe.

Die KSE-G kann prinzipiell in unterschiedlichen Erhebungsmodi eingesetzt werden. Im Rahmen unserer Validierung haben wir die Skala im CASI-Modus (Computer Assisted Self Interview) und im CAWI-Modus (Computer Assisted Web Interview) eingesetzt. Vor einem Einsatz der KSE-G in Mixed-Mode-Designs sollte allerdings eine Prüfung der Invarianz erfolgen. Eine Vorlage für die Erstellung der Erhebungsunterlagen, z. B. Fragebogen, Listenheft oder CAPI-Programmierung, ist in Appendix B zu finden. Dort sind die Instruktionen, die Items und die Antwortskala der KSE-G aufgeführt. Da die KSE-G in der in Appendix B dargestellten Form validiert wurde, empfehlen wir diese so oder möglichst ähnlich in die Erhebungsunterlagen zu übernehmen.

Die Durchführungszeit einer Vorversion der KSE-G mit acht Items im CASI-Modus liegt zwischen einer und eineinhalb Minuten. In 75% der CASI-Interviews dauert die Durchführung mit der Acht-Item-Version 83 Sekunden oder weniger (Perzentil<sub>75</sub> = 83). Eine ähnliche Durchführungsdauer ist im CAWI-Modus zu erwarten. Für die finale Sechs-Items-Version der KSE-G liegt die Durchführungsdauer noch unterhalb dieser Angaben.

---

<sup>1</sup> Ausgenommen sind Personen, deren sprachliche oder kognitive Fähigkeiten oder deren Wahrnehmungsfähigkeiten, z. B. durch Seh- oder Hörschwäche, unzureichend sind, um die Items zu verstehen.

## 4 Entwicklung und Validierung

---

Um die Skala zu konstruieren wurde zunächst ein Pool von Items zusammengestellt, die gemäß dem Ansatz von Crowne und Marlowe (1960) gesellschaftlich gewünschte aber meist seltene Verhaltensweisen oder gesellschaftlich sanktionierte aber häufig auftretende Verhaltensweisen erfassen. Dazu wurden bestehende Skalen, zum Beispiel die Soziale-Erwünschtheits-Skala-17 (SES-17; Stöber, 1999) und die Social Desirability Scale von Crowne und Marlowe (SDS-CM; 1960) herangezogen und Items entnommen. Diese wurden in Expertenreviews sprachlich überarbeitet, um deren Verständlichkeit für die anvisierte Zielgruppe und die Inhaltsvalidität zu erhöhen. Anschließend wurde der Itempool in einer Serie von Studien quantitativ analysiert (Item- und Strukturanalysen) und Items wurden selektiert. Die Strukturanalysen ergaben eine zweifaktorielle Struktur mit zwei korrelierten Faktoren, die als „Übertreibung positiver Qualitäten“ und „Minimierung negativer Qualitäten“ interpretiert wurden (vgl. Roth, Snyder & Pace, 1986). In einem iterativen Prozess wurden psychometrisch ungeeignete Items (geringe Faktorladungen, geringe Varianz, hohe Schiefe etc.) ausgeschieden und neue Items entwickelt, hinzugefügt und der Pool erneut an einer Stichprobe geprüft. Ergebnis der mehrstufigen Itemselektion waren die sechs Items der KSE-G, die zwei unterschiedliche Aspekte des Gamma-Faktors des Konstrukts Soziale Erwünschtheit mit jeweils drei Items erfassen. Die KSE-G wurde anschließend in drei alters-, geschlechts-, und bildungsheterogenen Stichproben validiert. Im Folgenden werden diese Stichproben und die Umstände der Erhebung näher beschrieben. Die Charakteristika der Stichproben können Tabelle 2 entnommen werden.

Stichprobe 1 ist eine Quotenstichprobe, geschichtet nach Geschlecht, Alter und Bildung ( $N = 741$ ), die im Internet erhoben wurde (CAWI). Grundgesamtheit waren die Teilnehmer eines Online-Access-Pools im Alter von 18 Jahren oder älter, die in Deutschland leben. Der Onlinefragebogen enthielt, neben der KSE-G, einige soziodemographische Angaben und verschiedene Skalen zur Messung psychologischer Merkmale. Die Bearbeitung des Onlinefragebogens dauerte im Mittel 23 Minuten ( $SD = 8$ ).

Stichprobe 2 mit  $N = 1134$  Befragungspersonen ist eine Zufallsstichprobe, die repräsentativ für die Wohnbevölkerung in Deutschland über einem Alter von 18 Jahren ist. Sie wurde mithilfe des ADM-Stichprobensystem F2F (Random Route) der Arbeitsgemeinschaft deutscher Marktforschungsinstitute gezogen. Die Daten wurden im CAPI- und im CASI-Modus erhoben (Dauer:  $M = 43$ ,  $SD = 13$ ). Die Fragebogenbatterie beinhaltete neben der KSE-G (im CASI-Modus) umfangreiche soziodemographische Angaben, verschiedene psychologische Maße und einige sozialwissenschaftliche Validierungsmaße. Die Items zu den soziodemographischen Angaben wurden größtenteils den demographischen Standards des Statistischen Bundesamtes entnommen (2010). Für die Validierung kamen Standardinstrumente, zum Beispiel zur Erfassung der fünf Hauptdimensionen der Persönlichkeit nach dem Fünf-Faktoren-Modell (BFI-10, Rammstedt & John, 2007; Rammstedt, Kemper, Klein, Beierlein & Kovaleva, 2012) und Items zur Messung von Werten nach dem Modell von Schwartz (1992; Schwartz & Boehnke, 2004) und Items zur Erfassung von Devianz aus dem ALLBUS (im CASI-Modus; Allgemeine Bevölkerungsumfrage der Sozialwissenschaften) zum Einsatz. Die KSE-G wurde in zwei unterschiedlichen experimentellen Bedingungen vorgegeben (im CASI-Modus). Befragungspersonen wurden randomisiert auf die Bedingungen zugewiesen. In einer Bedingung wurde die übliche Instruktion verwendet; in der anderen wurde die Befragungsperson darauf aufmerksam gemacht, dass die Items der KSE-G Verhaltensweisen beschreiben, die zwar gerne gesehen, aber nicht häufig gezeigt werden bzw. nicht gerne gesehen, aber üblicherweise häufig gezeigt werden. Im Anschluss an diese Aufklärung wurde die Befragungsperson gebeten „auch jetzt wieder möglichst ehrlich“ zu antworten.

Bei Stichprobe 3 handelt es sich um eine anfallende Onlinestichprobe (CAWI). Studierende unterschiedlicher Fachbereiche der Humboldt-Universität zu Berlin wurden über eine Email-Verteilerliste auf die Studie aufmerksam gemacht. Der Fragebogen enthielt einige soziodemographische Angaben,

einen Leistungstest und zahlreiche deutsche Übersetzungen von Skalen zur Erfassung von sozial erwünschtem Antwortverhalten, zum Beispiel die SD-Skala von Edwards (Lück & Timaeus, 1969), das BIDR von Paulhus (Musch, Brockhaus & Bröder, 2002) und die Lügenskala (L-Skala) aus dem MMPI-2 (Hathaway, McKinley & Engel, 2000) sowie weitere psychologische Testverfahren. Die Bearbeitung des Onlinefragebogens dauerte im Mittel 40 Minuten ( $SD = 19$ ).

*Tabelle 2:* Charakteristika der drei Validierungsstichproben.

	Stichprobe 1	Stichprobe 2	Stichprobe 3
<i>Stichprobe</i>			
Umfang [N]	741	1134	939
Art	Quote	Zufall	anfallend
Modus	CAWI	CAPI/CASI	CAWI
<i>Zusammensetzung</i>			
Geschlecht [% Frauen]	51.8%	55.6%	74.0%
Alter [M(SD)]	48.3 (13.0)	53.3 (18.4)	25.0 (5.4)
<i>Bildung</i>			
≤ 9 Jahre	40.1%	37.2%	k.A.
10 Jahre	29.1%	37.0%	k.A.
≥ 11 Jahre	30.8%	25.8%	k.A.

*Anmerkung:* CAPI = Computer Assisted Personal Interview, CASI = Computer Assisted Self Interview, CAWI = Computer Assisted Web Interview.

Um die psychometrische Güte der konstruierten Skala zu überprüfen, wurden auf der Grundlage der oben beschriebenen Stichproben Kennwerte für die Reliabilität und verschiedene Aspekte der Validität berechnet (für Details zur Validierung von Persönlichkeitsskalen siehe Bühner, 2011; Lienert & Raatz, 1998).

## 5 Gütekriterien

---

### 5.1 Objektivität

Unter Objektivität wird der Grad verstanden, in dem eine Messung unabhängig vom Untersucher ist (vgl. Lienert & Raatz, 1998). Diese bezieht sich auf verschiedene Phasen einer Untersuchung: Durchführung, Auswertung und Interpretation. Im Falle eines Face-to-face-Interviews hängt die Durchführungsobjektivität von dem Interviewer ab, der die Daten erhebt. Sie ist gegeben, wenn dieser sich bei der Vorgabe der Skala an die genauen Instruktionen und den Wortlaut der Items hält. Bei entsprechend geschulten Interviewern ist die Durchführungsobjektivität üblicherweise gewährleistet (Rammstedt, 2010b). Auswertungsobjektivität betrifft die numerische und kategoriale Auswertung des Antwortverhaltens der Befragten nach festgelegten Regeln (vgl. Lienert & Raatz, 1998). Diese ist bei der KSE-G vollständig gegeben, da die Verrechnungsvorschriften der Werte auf den Items klar festgelegt sind und keinen Interpretationsspielraum zulassen (siehe Abschnitt 2.3 und Appendix D). Interpretationsobjektivität ist gegeben, wenn die aus den Befragungsergebnissen gezogenen Schlüsse über verschiedene Forscher vergleichbar sind. Zur Maximierung der Interpretationsobjektivität sollte das Wissen der Forscher über die Messintention der Skala und über die Interpretation der quantitativen Messwerte vergleichbar sein (Rammstedt, 2010b). Durch die Standardisierung der Auswertung und die Zuweisung eines numerischen Messwerts, der die Ausprägung der Befragungsperson in Soziale Erwünschtheit beschreibt (siehe Befunde zur Validität in Abschnitt 5.3), kann auch die Interpretationsobjektivität der KSE-G als gegeben angesehen werden.

### 5.2 Reliabilität

Unter der Reliabilität oder Messgenauigkeit einer Skala versteht man den Grad der Genauigkeit, mit dem ein bestimmtes Merkmal erfasst wird (vgl. Lienert & Raatz, 1998). Die Reliabilität der KSE-G wurde im Rahmen von Strukturgleichungsmodellen (SEM; Jöreskog, 1969) auf Grundlage der Ladungen und Fehlervarianzen aus den Messmodellen in den drei Stichproben geschätzt. Als Schätzer wurde der Koeffizient  $\omega$  von McDonald (1999, S. 90) verwendet. Der Koeffizient gibt das Ausmaß an, in dem eine latente Variable (Konstrukt) von den Items geteilte Varianz reflektiert (Krohne & Hock, 2007). Laut Schweizer (2011) ist dieser Schätzer der Reliabilität besser geeignet als Cronbach  $\alpha$ . Die Interpretation der Höhe von McDonald  $\omega$  ist analog zu Cronbach  $\alpha$ . Die Schätzer der Reliabilität für die beiden Skalen betragen in Stichprobe 2  $\omega = .71$  für PQ+ und  $\omega = .78$  für NQ-. Dies entspricht einer für Gruppenuntersuchungen ausreichenden Reliabilität.

### 5.3 Validität

#### Inhaltliche Validität

Eine Skala ist inhaltlich valide, wenn ein Item das zu messende Konstrukt wirklich bzw. hinreichend präzise abbildet (Bühner, 2011). Eine empirische Prüfung der Inhaltsvalidität ist meist nicht möglich. Um zu überprüfen, ob ein Verfahren inhaltlich valide ist, wird üblicherweise dessen Konstruktion detailliert beleuchtet (Rammstedt, 2010b). Die inhaltliche Validität ist gegeben, wenn das zu messende Konstrukt a priori hinreichend definiert und die Items von einem Expertengremium im Hinblick auf ihre Gültigkeit beurteilt werden. Die inhaltliche Validität der KSE-G wurde sichergestellt, indem die Formulierung der Items eng an der Definition des Konstrukts orientiert war. Außerdem wurden die Itemformulierungen von einem Expertengremium generiert bzw. modifiziert und anschließend in einem kognitiven Pretest für die anvisierte Zielgruppe optimiert.

### Faktorielle Validität

Faktorielle Validität kann als gegeben angesehen werden, wenn die Annahmen über die dimensionale Struktur des zu erfassenden Konstrukts überprüfbar sind und belegt werden können. Im Rahmen einer CFA wurde geprüft, ob ein zweifaktorielles obliques Modell auf die Daten von Stichprobe 2 passt, da sich die Items der KSE-G in den explorativen Analysen mit den Konstruktionsstichproben meist auf zwei Faktoren aufteilen. Die Fitindizes zeigen eine akzeptable Passung des Modells auf die Daten an ( $\chi^2 = 20.7$ ,  $df = 8$ ,  $p < .05$ , RMSEA = .05, CFI = .99). Alle Items laden substantiell und hoch auf jeweils einem Faktor (siehe Tabelle 3). Die Interkorrelation der beiden Faktoren beträgt  $r = -.60$ .

Tabelle 3: Faktorladungen der KSE-G-Items in Stichprobe 2.

	F1	F2
(1) Es ist schon mal vorgekommen, dass ich jemanden ausgenutzt habe.		.78
(2) Auch wenn ich selbst gestresst bin, behandle ich andere immer freundlich und zuvorkommend.	.72	
(3) Manchmal helfe ich jemandem nur, wenn ich eine Gegenleistung erwarten kann.		.66
(4) Im Streit bleibe ich stets sachlich und objektiv.	.61	
(5) Ich habe schon mal Müll einfach in die Landschaft oder auf die Straße geworfen.		.74
(6) Wenn ich mich mit jemandem unterhalte, höre ich ihm immer aufmerksam zu.	.64	

Anmerkung:  $N = 566$  (Bedingung mit normaler Instruktion der KSE-G).

### Konstruktvalidität

Konstruktvalidität ist gegeben, wenn sich der Messwert einer Skala als Indikator für das Merkmal eignet, das mit der Skala gemessen werden soll. Die Eignung wird daran beurteilt, wie gut sich die Skala im Hinblick auf Hypothesen, die sich aus dem Modell des zu erfassenden Merkmals ableiten lassen, empirisch bewährt. Bei der Konstruktvalidierung werden anhand theoretischer Überlegungen positive, negative und Nullbeziehungen einer zu validierenden Skala mit anderen empirischen Indikatoren postuliert (nomologisches Netzwerk) und getestet, ob sich das vorhergesagte Muster empirisch belegen lässt (Krohne & Hock, 2007). Im Rahmen der Validierung der KSE-G wurde zunächst eine Korrelation mit alternativen Maßen der sozialen Erwünschtheit berechnet. Anschließend wurde versucht, weitere aus der Fachliteratur bekannte typische Korrelate von Gamma mit der KSE-G zu replizieren. Die praktische Bedeutsamkeit der im Folgenden berichteten empirisch ermittelten Validitätskoeffizienten aus den Stichproben 2 und 3 wird nach den Richtlinien von Cohen (1992) vorgenommen: kleiner Effekt ( $r = .10$ ), mittlere Effekt ( $r = .30$ ), starker Effekt ( $r = .50$ ).

Die höchste Korrelation weist die KSE-G erwartungsgemäß mit einem alternativen Maß für das Konstrukt auf, mit der „Lügenskala“ aus dem MMPI-2 (siehe Tabelle 4). Die „L-Skala“ misst „die Neigung, sich in der Testsituation zu verstellen und marginale Charakterschwächen zu leugnen, um als ideale Persönlichkeit zu erscheinen“ (Hank & Schwenkmezger, 2003, S. 295). Sowohl PQ+ als auch NQ- korrelieren moderat bis hoch mit der L-Skala. Geringe bis mittlere Korrelationen finden sich mit den Selbsttäuschungsskalen von Edwards und Paulhus. Unkorreliert ist die KSE-G mit der Impression Management-Skala (Fremdtäuschungsskala) aus dem BIDR. Neben den Korrelationen mit alternativen Maßen für soziale Erwünschtheit, stützen auch die Korrelationen mit den Devianzmaßen die Gültigkeit der KSE-G. Diese zeigen, dass Personen, die Gesetzesübertretungen wie Fahren ohne Fahrkarte oder Diebstahl zugeben, eher weniger dazu neigen ihre positiven Qualitäten zu übertreiben bzw. negative zu

untertreiben. Sowohl für PQ+ als auch für NQ- finden sich geringe bis mittlere negative Zusammenhänge mit den Devianzmaßen.

Tabelle 4: Validitätskoeffizienten der KSE-G.

		PQ+	NQ-
Soziodemographische Variablen <sup>1</sup>	<i>Alter</i>	.14**	.30**
	<i>Geschlecht</i>	.00	.18**
	<i>Bildung (Jahre)</i>	.00	-.10*
	<i>Bildung (Bücher)</i>	.00	-.03
	<i>Einkommen</i>	-.02	-.06
Devianz <sup>1</sup>	<i>Fahren ohne Fahrkarte</i>	-.22**	-.38**
	<i>Fahren mit über 0.5 ‰</i>	-.09*	-.29**
	<i>Diebstahl</i>	-.12**	-.33**
	<i>Steuerhinterziehung</i>	-.08	-.28**
Persönlichkeit <sup>1</sup>	<i>Neurotizismus</i>	-.18**	-.09*
	<i>Extraversion</i>	.06	.01
	<i>Offenheit</i>	.10*	.12**
	<i>Verträglichkeit</i>	.20**	.21**
	<i>Gewissenhaftigkeit</i>	.33**	.40**
Soziale Erwünschtheit <sup>2</sup>	<i>MMPI-2 Lügenskala</i>	.43**	-.38**
	<i>BIDR-SDS</i>	.12**	-.03
	<i>BIDR-IM</i>	.05	.00
	<i>SDS-E</i>	.11**	-.25**

Anmerkung: PQ+ = Übertreibung positiver Qualitäten, NQ- = Minimierung negativer Qualitäten, MMPI-2 = Minnesota Multiphasic Personality Inventory-2, BIDR-SDS/IM = Balanced Inventory of Desirable Responding - Subscale Self-Deception/Impression Management, SDS-E = Edwards Social Desirability Scale. Kodierung Geschlecht 1 = männlich, 2 = weiblich. <sup>1</sup> Stichprobe 2 (N = 566), <sup>2</sup> Stichprobe 3 (N = 309-939). \*  $p < .05$ , \*\*  $p < .01$ .

Mit soziodemographischen Variablen zeigen sich differentielle Zusammenhänge für die KSE-G. PQ+ und NQ- sind geringfügig bis moderat mit dem Alter korreliert. Demnach neigen eher ältere Menschen zur Übertreibung positiver Qualitäten und zur Untertreibung negativer. Geschlecht ist ausschließlich mit der Untertreibung negativer Qualitäten assoziiert. Dies gilt der Richtung des Zusammenhangs nach eher für Frauen als für Männer.

Anhand der folgenden Korrelationen kann festgestellt werden, ob bei der Beantwortung von Persönlichkeitsitems in Stichprobe 2 ein Bias auftrat und welcher Art dieser Bias war. Dazu wurden Korrelationen der KSE-G mit den Items aus dem BFI-10 und den Items zur Erfassung von Werten nach Schwartz (1992; Schwartz & Boehnke, 2004) berechnet. Substantielle geringe bis mittlere Korrelationen weist die KSE-G mit allen Skalen des BFI-10 außer Extraversion auf. Die höchsten Korrelationen finden sich für Gewissenhaftigkeit und Verträglichkeit. Auch mit Neurotizismus korreliert die KSE-G (negativ). Dieses Muster an Befunden spricht für einen moralistischen Bias (assoziiert mit dem Gamma-Faktor) in den Antworten der Befragungspersonen (vgl. Paulhus, 2002).

Die Items zur Erfassung von Werten nach dem Modell von Schwartz wurden auf der Grundlage der Konstruktbeschreibung von Bakan (1966) und Paulhus (2002) ad hoc in Agency- und Communion-Werte klassifiziert. Die Korrelationen zwischen der KSE-G und den Werteitems sind in Tabelle 5 dargestellt. Die Werteitems, die sich auf Communion-Werte beziehen, korrelieren durchweg substantiell in

geringer bis moderater Höhe mit der KSE-G. Die über die Items aggregierte Korrelation beträgt .21 für PQ+ und .25 für NQ-. Für die Agency-Items zeigen sich nur marginal positive bzw. einige gering bis moderat negative Korrelationen. Die über die Items aggregierte Korrelation beträgt -.03 für PQ+ und -.11 für NQ-. Auch dieses Befundmuster legt einen moralistischen Bias Nahe, der sich in den Antworten der Befragungspersonen auf die Werteeitems manifestiert hat.

*Tabelle 5:* Korrelationen der KSE-G mit Werteeitems aus dem Modell von Schwartz (1992; Schwartz & Boehnke, 2004) in Stichprobe 2.

	PQ+	NQ-
<i>Communion-Werte</i>		
Alle Gesetze zu befolgen, ist ihm/ihr wichtig.	.23**	.27**
Es ist ihm/ihr sehr wichtig, den Menschen zu helfen, die ihm/ihr am Herzen liegen.	.21**	.27**
Es ist ihm/ihr wichtig, traditionelle Werte und Glaubensvorstellungen zu bewahren.	.17**	.20**
Seiner/Ihrer Meinung nach ist es wichtig, dass jeder Mensch auf der Welt die gleichen Chancen hat im Leben.	.21**	.22**
Er/Sie ist der festen Überzeugung, dass er/sie sich für die Umwelt einsetzen soll.	.22**	.27**
<i>Agency-Werte</i>		
Er/Sie möchte, dass die Leute seine/ihre Leistungen bewundern.	-.05	-.24**
Er/Sie möchte, dass die Leute tun was er/sie sagt.	-.08*	-.22**
Reich zu sein, ist ihm/ihr wichtig.	-.15**	-.26**
Es ist ihm/ihr wichtig niemals ein Risiko einzugehen.	.05	.15**
Die Freiheit zu haben, zu wählen, was er/sie tut, ist ihm/ihr wichtig.	.08**	.02

*Anmerkung:* PQ+ = Übertreibung positiver Qualitäten, NQ- = Untertreibung negativer Qualitäten.  $N = 566$  (Bedingung mit normaler Instruktion der KSE-G).

Als weiterer Aspekt der Konstruktvalidität neben den oben berichteten Validitätskoeffizienten wurde geprüft, ob die Skalenwerte für PQ+ und NQ- änderungssensitiv sind, d.h. durch experimentelle Manipulation verändert werden können, da Paulhus in seinem integrativen Modell auch einen situationsabhängigen Aspekt des Gamma-Faktors konzeptualisiert: Communion-Management. Die experimentelle Variation in Stichprobe 2, die Befragungspersonen um ehrliche Antworten zu bitten, führte zu differentiellen Effekten auf die KSE-G. Für PQ+ fanden sich keine Unterschiede zwischen den experimentellen Bedingungen ( $t = 1.43$ ,  $df = 1132$ ,  $p > .05$ , Hedges  $g = .06$ ). Für NQ- zeigten sich hingegen substantielle und praktisch bedeutsame Unterschiede ( $t = 2.11$ ,  $df = 1132$ ,  $p < .05$ , Hedges  $g = .13$ ). Befragungspersonen, die gebeten wurden möglichst ehrlich zu antworten, zeigten geringere Werte bei der Untertreibung negativer Qualitäten ( $M = 3.14$ ,  $SD = .81$ ) im Vergleich zu Befragungspersonen, die nur die übliche Instruktion für die KSE-G bekamen ( $M = 3.24$ ,  $SD = .78$ ).

## 5.4 Referenzwerte

Im Appendix A sind Referenzwerte in Form von Gruppenmittelwerten und Standardabweichungen für die KSE-G abgedruckt. Diese wurden anhand der Zufallsstichprobe (Stichprobe 2) ermittelt und erlau-

ben dem Anwender einen Vergleich der KSE-G-Werte aus seiner Untersuchung mit denen relevanter Subgruppen aus einer bevölkerungsrepräsentativen Zufallsstichprobe, zum Beispiel von Männern oder Frauen, von Personen mit unterschiedlicher Schulbildung oder unterschiedlichen Alters. Die Altersgruppen in Tabelle A wurden den Lebensphasen der bundesdeutschen Gesellschaft angepasst. Die Zeit von 18 bis 35 Jahren ist die der beruflichen Ausbildung und Familiengründung. Die Zeit der beruflichen Festigung, Karriere, Betreuung heranwachsender Kinder und Pflege älterer Angehöriger fällt in die Zeit zwischen 36 und 65 Jahren. Die dritte Lebensphase beginnt im Alter von 65 Jahren, wenn die berufliche Tätigkeit in den meisten Fällen abgeschlossen ist. Die Aufteilung der Bildungsstufen wurde nach der Dauer der schulischen Allgemeinbildung vorgenommen. Dabei gilt die Dauer der schulischen Bildung bis einschließlich 9 Jahren als geringes Bildungsniveau. Bei einer Schuldauer von 10 oder 11 Jahren handelt es sich um ein mittleres Bildungsniveau und bei mehr als 11 Jahren um ein hohes Bildungsniveau.

## 6 Kurzfassung

---

### Diagnostische Zielsetzung

Die Kurzsкала Soziale Erwünschtheit-Gamma (KSE-G) wurde entwickelt, um eine ökonomische Messung eines Teilaspekts des psychologischen Merkmals soziale Erwünschtheit, den Gamma-Faktor, in sozialwissenschaftlichen Untersuchungen zu ermöglichen.

### Aufbau

Die KSE-G erfasst zwei Aspekte des Gamma-Faktors von sozialer Erwünschtheit, die Übertreibung positiver Qualitäten (PQ+) und die Minimierung negativer Qualitäten (NQ-) mit jeweils drei Items auf einer fünfstufigen Ratingskala von „trifft gar nicht zu“ (0) bis „trifft voll und ganz zu“ (4).

### Grundlagen und Konstruktion

Der KSE-G liegt eine in der psychologischen Forschung weit verbreitete Konstruktdefinition von sozialer Erwünschtheit zugrunde, gemäß der soziale Erwünschtheit als Tendenz verstanden wird überwiegend positive Beschreibungen der eigenen Person abzugeben. Die Kurzsкала wurde auf der Grundlage eines integrativen mehrdimensionalen Modells des Konstrukts entwickelt (Paulhus, 2002). Die Items der KSE-G wurden bereits bestehenden Verfahren entnommen und verbessert. Sie erfassen gesellschaftlich gewünschte aber meist seltene Verhaltensweisen oder gesellschaftlich sanktionierte aber häufig auftretende Verhaltensweisen. Die Items wurden in Expertenreviews und kognitiven Pretests sprachlich überarbeitet und anschließend mehreren umfangreichen alters-, geschlechts- und bildungs-heterogenen Stichproben vorgegeben, um die psychometrischen Gütekriterien zu ermitteln.

### Gütekriterien

*Reliabilität* Die Reliabilität der KSE-G wurde anhand des Koeffizienten  $\omega$  von McDonald (1999, S. 90) ermittelt. Die Interpretation der Höhe von McDonald  $\omega$  entspricht der Interpretation des bekannteren, aber weniger geeigneten Cronbach  $\alpha$ . Die Schätzer der Reliabilität liegen bei  $\omega = .71$  für PQ+ und  $\omega = .78$  für NQ-. Dies entspricht einer für Gruppenuntersuchungen ausreichenden Reliabilität (vgl. Aiken & Groth-Marnat, 2006).

*Validität* Anhand von drei Stichproben wurde die internale Struktur der KSE-G überprüft, Validitätskoeffizienten mit diversen Maßen berechnet und Mittelwertunterschiede auf der KSE-G nach experimenteller Manipulation betrachtet, um die Konstruktvalidität der Kurzsкала zu prüfen. Die Faktoranalysen zeigen, dass der Gamma-Faktor zwei Aspekte, die Übertreibung positiver und Untertreibung negativer Qualitäten beinhaltet (vgl. Roth, Snyder & Pace, 1986). Korrelative Analysen ergaben, dass die KSE-G mit alternativen Maßen für das Konstrukt korreliert. In den Antworten der Befragungspersonen auf Persönlichkeits- und Werteitems wurde ein moralistischer Bias evident, der in der Fachliteratur mit dem Gamma-Faktor assoziiert wird. Schließlich konnte anhand einer experimentellen Manipulation der Motivation der Befragungspersonen die Änderungssensitivität der KSE-G belegt werden. Die durchgeführten Analysen zeigen, dass die KSE-G eine valide ökonomische Erfassung des Gamma-Faktors sozial erwünschten Antwortverhaltens erlaubt.

## 7 Bewertung

---

Die KSE-G ist eine ökonomische Skala zur Erfassung des psychologischen Merkmals soziale Erwünschtheit (Gamma-Faktor) in sozialwissenschaftlichen Untersuchungen, die starken zeitlichen und monetären Restriktionen unterliegen. Die Skala ist einfach in verschiedenen Erhebungsmodi zu administrieren. Validitätsbelege wurden für den CASI-Modus (Selbstaussfüller) und den CAWI-Modus (Onlinefragebogen) erbracht. Die Invarianz zwischen den Erhebungsmodi wurde bisher allerdings noch nicht geprüft. Ob Korrelationen und Mittelwerte der KSE-G daher über Stichproben hinweg vergleichbar sind, die in unterschiedlichen Modi erhoben wurden, kann zurzeit nicht vorausgesetzt werden. Vor dem Einsatz der KSE-G in Mixed-Mode-Designs sollte dieser Beleg erbracht werden (für aktualisierte Informationen bezüglich der psychometrischen Güte siehe [www.gesis.org/kurzskalen-psychologischer-merkmale](http://www.gesis.org/kurzskalen-psychologischer-merkmale)). Die Erhebung von Daten mit der KSE-G dauert deutlich weniger als eine Minute. Die Auswertung erfolgt streng standardisiert.

Die Überprüfung der Messgenauigkeit der KSE-G ergab eine für Gruppenuntersuchungen ausreichenden Reliabilität der beiden Skalen PQ+ und NQ-. Die Analysen zur internalen Struktur der KSE-G zeigten Ergebnisse, die auf den ersten Blick als inkompatibel mit dem integrativen Modell von Paulhus (2002) schienen. Er postuliert, dass Gamma einen Selbsttäuschungsstil und einen Fremdtäuschungsstil beinhaltet. Faktoranalytisch wurden in der vorliegenden Arbeit allerdings die beiden Faktoren Übertreibung positiver Qualitäten und Untertreibung negativer Qualitäten gefunden, die auch Roth, Snyder und Pace (1986) berichten. Eine Vermutung diesbezüglich ist, dass Gamma in Abhängigkeit von motivationalen Anreizen der jeweiligen Testsituation in unterschiedliche Faktoren „zerfällt“, entweder in die von Paulhus postulierten Selbsttäuschungsstil und Fremdtäuschungsstil oder die von Roth et al. gefundenen Faktoren Übertreibung positiver Qualitäten und Untertreibung negativer Qualitäten. Für diese Vermutung spricht, dass die von Paulhus postulierten Faktoren unterschiedlich änderungssensitiv sind. Insbesondere der Fremdtäuschungsstil Communion-Management beinhaltet bewusste Übertreibung des positiven Eindrucks, z.B. bei Anwesenheit wichtiger Personen bzw. Publikum. Die Binnenstruktur von Gamma könnte demnach bei starker Aktivierung von Communion-Werten durch situative Gegebenheiten aussehen, wie von Paulhus beschrieben. In der vorliegenden Arbeit wurde jedenfalls die Faktoren Übertreibung positiver Qualitäten und Untertreibung negativer Qualitäten explorativ gefunden und per CFA gesichert. Anhand der Validitätskoeffizienten konnte gesichert werden, dass die KSE-G Gamma erfasst, da (1) die KSE-G mit alternativen Maßen für das Konstrukt korreliert, insbesondere der L-Skala aus dem MMPI-2, und (2) in den Antworten der Befragungspersonen auf Persönlichkeits- und Wertetems ein deutlicher moralistischer Bias gefunden wurde, der in der Fachliteratur mit dem Gamma-Faktor assoziiert wird. Die Befunde zeigen allerdings auch, dass eine Unterteilung in PQ+ und NQ- vertretbar ist, weil (1) differentielle Korrelationen der beiden Skalen auftraten, insbesondere mit dem soziodemographischen Kriterien Alter, Geschlecht und Bildung, und (2) die milde experimentelle Manipulation einen geringen aber substantiellen Effekt auf Untertreibung negativer Qualitäten, nicht aber auf Übertreibung positiver Qualitäten hatte.

Die empirischen Belege der Validierungsstudien sprechen dafür, dass die KSE-G nicht nur eine ökonomische, sondern auch eine reliable und valide Erfassung des Gamma-Faktors sozial erwünschten Antwortverhaltens erlaubt. Einschränkend sei allerdings an dieser Stelle erwähnt und betont, dass die KSE-G für eine statistische Kontrolle der bei einer Messung von Persönlichkeit oder Einstellungen auftretenden systematischen Fehlervarianz nicht geeignet ist. Lange Jahre wurde in der Fachliteratur propagiert, dass Skalen zur Erfassung von sozialer Erwünschtheit absichtliche Antwortverzerrungen (oder „Lügen“) reliabel messen können und sich diese Maße daher eigenen, nicht-intendierte Varianz in Messungen zu kontrollieren. Dieses Vorgehen ist problematisch, weil Antwortverzerrungen weitgehend eine Funktion substantieller Persönlichkeitsmerkmale sind. Dies haben auch die Befunde der vorliegenden Arbeit anhand des moralistischen Bias gezeigt. Mittlerweile wird in der Fachliteratur die Meinung

vertreten, dass klassische Soziale-Erwünschtheits-Skalen diesem Anspruch nicht genügen können, sich neue alternative Maße für Antwortverzerrungen wie beispielsweise der Wortschatz und Overclaiming-Test (WOCT; Ziegler, Kemper & Rammstedt, in press) bzw. das Overclaiming-Konzept (Paulhus, Harms, Bruce & Lysy, 2003) allerdings dafür eignen könnten (für eine Übersicht solcher Befunde siehe Bing, Kluemper, Davison, Taylor & Novicevic, 2011).

## 8 Literatur

---

- Aiken, L. R. & Groth-Marnat, G. (2006). *Psychological Testing and Assessment*. Boston: Pearson.
- Allison, P. J., Guichard, C., Fung, K & Gilain, L. (2003). Dispositional optimism predicts survival status 1 year after diagnosis in head and neck cancer patients. *Journal of Clinical Oncology*, 21, 543-548.
- Arthur, W., Jr. & Graziano, W. G. (1996). The five-factor model, conscientiousness, and driving accident involvement. *Journal of Personality*, 64, 593-618.
- Bakan, D. (1966). *The duality of human existence. An essay on psychology and religion*. Chicago: Rand McNally.
- Barrick, M. R. & Mount, M. K. (1991). The big five personality dimensions and job performance: a meta-analysis. *Personnel Psychology*, 44, 1-26.
- Bing, M. N., Kluemper, D., Kristl Davison, H., Taylor, S., & Novicevic, M. (2011). Overclaiming as a measure of faking. *Organizational Behavior and Human Decision Processes*, 116, 148-162.
- Block, J., Block, J. H. & Keyes, S. (1988). Longitudinally foretelling drug usage in adolescence: Early childhood personality and environmental precursors. *Child Development*, 59, 336-355.
- Block, J., Gjerde, P. F. & Block, J. H. (1991). Personality antecedents of depressive tendencies in 18-year-olds: A prospective study. *Journal of Personality and Social Psychology*, 60, 726-738.
- Borghans, L., Duckworth, A. L., Heckman, J. J. & ter Weel, B. (2008). *The Economics and Psychology of Personality Traits. IZA Discussion Paper No. 3333*. Bonn: IZA.
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion. PS Psychologie*. München: Pearson Studium.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159.
- Crowne, D. P. & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24, 349-354.
- Edwards, A. L. (1957). *The social desirability variable in personality assessment and research*. Westport: Greenwood Press.
- Goldberg, L. R. (2005). *Why Personality Measures Should Be Included in Epidemiological Surveys: A brief Commentary and a Reading List*. (Oregon Research Institute, Hrsg.).
- Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence*, 24, 79-132.
- Gottfredson, L. S. & Deary, I. J. (2004). Intelligence predicts health and longevity, but why? *Current Directions in Psychological Science*, 13, 1-4.
- Hank, P. & Schwenkmezger, P. (2003). Unvertretbar nach 40 Jahren der Anwendung - Meinungen über MMPI-2 gehen weit auseinander. *Report Psychologie*, 28, 295-303.
- Hathaway, S. R. & McKinley, J. C. & Engel, R. (2000) *Minnesota Multiphasic Personality Inventory 2*. Göttingen: Hogrefe.
- International Test Commission (2010). *International Test Commission Guidelines for Translating and Adapting Tests*. Gefunden am 01.02.2012 unter <http://www.intestcom.org>.
- Jakoby, N. & Jacob, R. (1999). Messung von internen und externen Kontrollüberzeugungen. *ZUMA-Nachrichten*, 45, 61-71.

- Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, *34*, 183-202.
- Kemper, C. J., Beierlein, C., Kovaleva, A., & Rammstedt, B. (in Druck). Entwicklung und Validierung einer ultrakurzen Operationalisierung des Konstrukts Optimismus-Pessimismus – Die Skala Optimismus-Pessimismus-2 (SOP2). *Diagnostica*.
- Krohne, H. W. & Hock, M. (2007). *Psychologische Diagnostik: Grundlagen und Anwendungsfelder*. Stuttgart: Kohlhammer.
- Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID). (Hrsg.). (2011). *Verzeichnis Testverfahren. Kurznamen. Langnamen. Autoren. Testrezensionen* (18., aktualisierte Auflage). Trier: ZPID [Online verfügbar unter: [http://www.zpid.de/pub/tests/verz\\_teil1.pdf](http://www.zpid.de/pub/tests/verz_teil1.pdf); Stand: 30.6.2011].
- Lienert, G. A. & Raatz, U. (1998). Testaufbau und Testanalyse (6. Aufl.). *Grundlagen Psychologie*. Weinheim: Julius Beltz.
- Lück, H. E. & Timaeus, E. (1969). Skalen zur Messung Manifeste Angst (MAS) und sozialer Wünschbarkeit (SDS-E und SDS-CM). *Diagnostica*, *15*, 134-141.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah: Erlbaum.
- Musch, J., Brockhaus, R., & Bröder, A. (2002). Ein Inventar zur Erfassung von zwei Faktoren sozialer Erwünschtheit. *Diagnostica*, *48*, 121-129.
- Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17-59). New York: Academic Press.
- Paulhus, D. L. (2002). Social Desirable Responding. The Evolution of a Construct. In H. I. Braun & D. N. Jackson (Hrsg.), *The role of constructs in psychological and educational measurement*. Mahwah: Erlbaum.
- Paulhus, D. L., Harms, P. D., Bruce, M. N. & Lysy, D. C. (2003). The Over-Claiming Technique. Measuring Self-Enhancement Independent of Ability. *Journal of Personality and Social Psychology*, *84*(4), 890-904.
- Rammstedt, B. (2010a). Subjective indicators. In *Building on progress. Expanding the research infrastructure for the social, economic, and behavioral sciences* (S. 813-824). Opladen: Budrich Uni-Press.
- Rammstedt, B. (2010b). Reliabilität, Validität, Objektivität. In Wolf, C. & Best, H. (Hrsg.). *Handbuch der sozialwissenschaftlichen Datenanalyse* (S. 239-258). Wiesbaden: VS.
- Rammstedt, B. & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, *41*, 203-212.
- Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C., & Kovaleva, A. (2012). Eine kurze Skala zur Messung der fünf Dimensionen der Persönlichkeit: Big-Five-Inventory-10 (BFI-10) (GESIS Working Papers 2012|23). Köln: GESIS.
- Rasmussen, H. N., Scheier, M. F. & Greenhouse, J. B. (2009). Optimism and physical health: A meta-analytic review. *Annals of Behavioral Medicine*, *37*, 239-256.
- Rat für Sozial- und Wirtschaftsdaten (2010). *Building on progress: Expanding the research infrastructure for the social, economic, and behavioral sciences*. Opladen: Budrich Uni-Press.

- Roth, D. L., Snyder, C. R., & Pace, L. M. (1986). Dimensions of favorable self-presentation. *Journal of Personality and Social Psychology*, 51(4), 867–874. doi:10.1037/0022-3514.51.4.867
- Schmidt, F. L. & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262–274.
- Schumann, S. & Schoen, H. (2005). *Persönlichkeit: Eine vergessene Größe der empirischen Sozialforschung*. Wiesbaden: VS Verlag.
- Schweizer, K. (2011). On the changing role of Cronbach  $\alpha$  in the evaluation of the quality of a measure. *European Journal of Psychological Assessment*, 27, 143–144.
- Schwartz, S. H. (1992). Universals in the content and structure of values: Theory and empirical tests in 20 countries. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 25) (pp. 1–65). New York: Academic Press.
- Schwartz, S. H. & Boehnke, K. (2004). Evaluating the structure of human values with confirmatory factor analysis. *Journal of Research in Personality*, 38, 230–255.
- Stanton, J. M., Sinar, E. F., Balzer, W. K. & Smith, P. C. (2002). Issues and strategies for reducing the length of self-report scales. *Personnel Psychology*, 55, 167–194.
- Stöber, J. (1999). Die Soziale Erwünschtheits-Skala-17 (SES-17): Entwicklung und erste Befunde zu Reliabilität und Validität. *Diagnostica*, 45, 173–177.
- Strenze, T. (2007). Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence*, 35, 401–426.
- Ziegler, M., Kemper, C. J., & Rammstedt, B. (in press). The Vocabulary and Overclaiming Test (VOC-T). *Journal of Individual Differences*.

## Appendix A

### Referenzwerte

Table A1: KSE-G PQ+ (Übertreibung positiver Qualitäten)

Geschlecht	Bildung	Altersgruppen						Gesamt (Alter)	
		18-35		36-65		>65		M	SD
		M	SD	M	SD	M	SD		
Männlich	gering	2.33	0.77	2.72	0.72	2.84	0.67	2.73	0.72
	mittel	2.51	0.76	2.73	0.71	2.90	0.72	2.74	0.73
	hoch	2.70	0.82	2.74	0.66	2.89	0.74	2.76	0.73
	Gesamt	2.54	0.80	2.73	0.70	2.87	0.69	2.74	0.72
Weiblich	gering	2.21	0.72	2.53	0.73	2.79	0.69	2.64	0.73
	mittel	2.79	0.70	2.66	0.71	2.95	0.84	2.74	0.78
	hoch	2.64	0.81	2.86	0.66	3.26	0.49	2.83	0.76
	Gesamt	2.65	0.76	2.67	0.70	2.88	0.72	2.73	0.76
Gesamt (Geschlecht)	gering	2.28	0.75	2.62	0.73	2.82	0.68	2.68	0.73
	mittel	2.71	0.72	2.68	0.76	2.92	0.77	2.74	0.76
	hoch	2.66	0.81	2.80	0.71	3.04	0.67	2.80	0.75
	Gesamt	2.61	0.76	2.70	0.74	2.88	0.71	2.73	0.74

Table A2: KSE-G NQ- (Untertreibung negativer Qualitäten)

Geschlecht	Bildung	Altersgruppen						Gesamt (Alter)	
		18-35		36-65		>65		M	SD
		M	SD	M	SD	M	SD		
Männlich	gering	2.51	1.08	3.08	0.84	3.36	0.66	3.15	0.83
	mittel	2.61	1.01	3.08	0.72	3.43	0.62	3.09	0.80
	hoch	2.72	0.93	2.96	0.96	3.20	0.84	2.94	0.94
	Gesamt	2.63	0.99	3.04	0.84	3.35	0.68	3.07	0.85
Weiblich	gering	2.58	0.78	3.26	0.74	3.56	0.67	3.36	0.76
	mittel	2.99	0.79	3.24	0.71	3.59	0.62	3.23	0.74
	hoch	3.04	0.77	3.40	0.59	3.38	0.70	3.26	0.70
	Gesamt	2.96	0.79	3.29	0.69	3.55	0.66	3.28	0.74
Gesamt (Geschlecht)	gering	2.54	0.95	3.17	0.79	3.46	0.67	3.26	0.80
	mittel	2.88	0.87	3.18	0.72	3.50	0.62	3.18	0.77
	hoch	2.91	0.85	3.19	0.82	3.28	0.78	3.11	0.83
	Gesamt	2.83	0.89	3.18	0.77	3.45	0.68	3.19	0.80

Anmerkungen. N = 566.

## Appendix B

### KSE-G

Die folgenden Aussagen können auf Sie selbst mehr oder weniger zutreffen. Bitte geben Sie bei jeder Aussage an, wie sehr die Aussage auf Sie zutrifft.

	trifft gar nicht zu	trifft wenig zu	trifft etwas zu	trifft ziemlich zu	trifft voll und ganz zu
(1) Es ist schon mal vorgekommen, dass ich jemanden ausgenutzt habe.	0	1	2	3	4
(2) Auch wenn ich selbst gestresst bin, behandle ich andere immer freundlich und zuvorkommend.	0	1	2	3	4
(3) Manchmal helfe ich jemandem nur, wenn ich eine Gegenleistung erwarten kann.	0	1	2	3	4
(4) Im Streit bleibe ich stets sachlich und objektiv.	0	1	2	3	4
(5) Ich habe schon mal Müll einfach in die Landschaft oder auf die Straße geworfen.	0	1	2	3	4
(6) Wenn ich mich mit jemandem unterhalte, höre ich ihm immer aufmerksam zu.	0	1	2	3	4

## Appendix C

### KSE-G (English version)

The following statements may apply more or less to you personally. Please indicate to what extent they apply to you.

	Doesn't apply at all	Applies a bit	Applies somewhat	Applies mostly	Applies completely
(1) It has happened that I have taken advantage of someone in the past.	0	1	2	3	4
(2) Even if I am feeling stressed, I am always friendly and polite to others.	0	1	2	3	4
(3) Sometimes I only help people if I expect to get something in return.	0	1	2	3	4
(4) In an argument, I always remain objective and stick to the facts.	0	1	2	3	4
(5) I have occasionally thrown litter away in the countryside or on to the road.	0	1	2	3	4
(6) When talking to someone I always listen carefully to what the other person says.	0	1	2	3	4

---

## Appendix D

---

### SPSS-Syntax zur Auswertung der KSE-G

```
COMPUTE PQplus=MEAN(item02,item04,item06).  
EXECUTE.
```

```
COMPUTE NQminus=MEAN(item01,item03,item05).  
EXECUTE.
```

#### VARIABLE LABELS

PQplus 'Übertreibung positiver Qualitäten'

NQminus 'Untertreibung negativer Qualitäten'