

CSI



UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Teilauswertung des DSEM 2023-2024 zur Wirkungsmessungspraxis in der Social Economy in Deutschland

In Kooperation mit:

Social Entrepreneurship Netzwerk Deutschland e. V.

c/o Impact Hub Rollbergstraße 28a

12053 Berlin

<mailto:info@send-ev.de>

<https://www.send-ev.de>

Autor:

Dr. Filip Zieliński

Forschungsstelle

Centrum für Soziale Investitionen und Innovationen (CSI)

Max-Weber-Institut für Soziologie

Bergheimer Str. 58

D-69115 Heidelberg

filip.zielinski@csi.uni-heidelberg.de

Heidelberg, Juni 2025

Empfohlene Zitierweise: Zieliński, F. (2025). *Teilauswertung des DSEM 2023-2024 zur Wirkungsmessungspraxis in der Social Economy in Deutschland*. Research Paper, Centrum für Soziale Investitionen und Innovationen, Universität Heidelberg. doi.org/10.11588/heidok.00036682

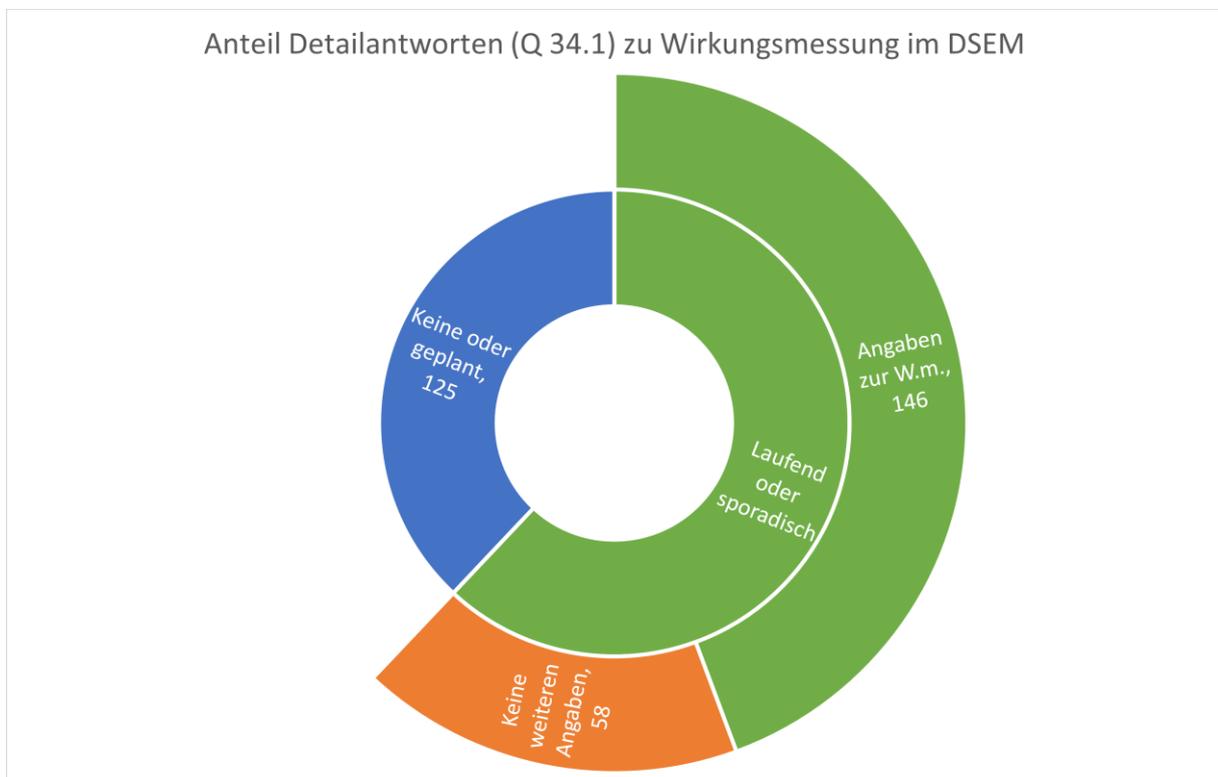
Auswertung

Datengrundlage der vorliegenden Analyse ist der Deutsche Social Entrepreneurship Monitor 2023-2024, der vom Social Entrepreneurship Netzwerk Deutschland e. V. (SEND) durchgeführt worden ist. Das CSI wurde von SEND eingeladen, als wissenschaftlicher Kooperationspartner einen Teil der DSEM-Umfrage auszuwerten. Für das CSI bietet diese Kooperation eine Gelegenheit zur Vertiefung und zum Transfer seiner laufenden Forschungsarbeiten zu Wirkungsmessung und -management (IMM).

In der Publikation der Ergebnisse des DSEM wurden die meisten Umfragedaten zum Thema Wirkungsmessung bereits durch SEND ausgewertet (siehe [DSEM 2023-2024](#), Seite 34-40). Die Umfrage beinhaltet allerdings ein Frageitem, das bisher nicht analysiert wurde, vor allem weil es sich um eine offene Frage handelt, die von den Befragten mittels eines Textfelds beantwortet werden konnte und die nur mit einem gewissen Aufwand ausgewertet werden kann:

“What tools, methods, frameworks, certifications or approaches does your organisation use to measure or evaluate its impact?” (Q 34.1)

Von den insgesamt 329 im DSEM Befragten Personen (bzw. Organisationen) gaben 125 (38 %) an, ihre soziale/ökologische Wirkung nicht zu messen, oder diese nicht zu messen aber Wirkungsmessung in Zukunft zu planen (siehe Diagramm unten). Die übrigen 204 Befragten (62 %) gaben an, ihre Wirkung zumindest sporadisch, oder aber regelmäßig, zu messen. Es stellt sich jedoch die Frage, was die Befragten unter „Wirkung messen“ verstanden haben und wie diese Wirkungsmessung jeweils gestaltet war. Die Antworten auf Frage Q 34.1 verspricht hierzu wertvolle Hinweise.



Zu bedenken ist hierbei, dass ein Teil der Befragten, die angab, Wirkung zu messen, keine weiteren Angaben im offenen Textfeld zu Q34.1 gemacht hat (58 Fälle). Die übrigen 146 Befragten haben das Feld ausgefüllt; diese Antworten wurden zur weiteren Analyse manuell erstellten Kategorien zugeordnet (z.B. „SRS“ oder „GWÖ“). Zahlreiche Antworten beinhalteten Hinweise auf mehr als ein Tool/Framework/Methode/Zertifikat, sodass die entsprechenden Fälle mehreren Kategorien zugeteilt wurden. Die Kategorien wurden induktiv anhand der Antworten erstellt, also im Laufe der Durchsicht der Daten (und nicht im Vorhinein) definiert und zusammengeführt. Die Kategorien lassen sich recht gut anhand der zusammenfassenden Kategorien gruppieren, die in der Frage formuliert werden (Tools, Frameworks, Methods, Certificates).

Eine vollständige Auswertung aller gefundenen Kategorien ist in Tabelle 2 zu finden. Hier zeigt sich, dass die Nennungen – zum größten Teil aber mit einzelnen Ausnahmen – durchaus dem in Evaluationsforschung und -praxis gängigen Begriff von Wirkung und Wirkungsmessung entsprechen. Möglicherweise hat der Hinweis auf die „sozialen/ökologischen Wirkungen“ zur augenscheinlichen Konstruktvalidität positiv beigetragen. Jedenfalls konnten die meisten Befragten relevante Antworten machen. Zu erwähnen ist allerdings, dass in 32 Fällen Nennungen gemacht wurden, die als „nicht spezifizierte Datenerhebungsmethoden“ zusammengefasst sind: das sind Nennungen, die nicht wirkungsmessungsrelevant sind oder zu vage um kategorisiert werden zu können. Recht oft wurden konkrete Indikatoren genannt, die zur Wirkungsmessung genutzt werden; in vielen Fällen dürften dies Indikatoren auf der Output-, nicht auf der Outcome oder Impact Ebene sein, was sich anhand der vorliegenden Daten aber nur grob einschätzen lässt.

Beschränkt man die Auswertung auf Nennungen, die von mindestens 5 Befragten vorgenommen wurden, ergibt sich folgendes Bild:

Tabelle 1

Auswertung der am häufigsten genannten Tools, Frameworks, Methods & Certificates				
	<i>lediglich Fälle mit mindestens 5 Nennungen; Mehrfachnennungen möglich</i>			
	Anzahl Nennungen	Anteil Nennungen	Anteil an Respondenten mit Wirkungsmessung	Anteil an Respondenten insgesamt
Tools & Frameworks				
Phineo/Wirkungstreppe	23	30%	11%	7%
GWÖ, Gemeinwohl-Bilanz, ECG	10	13%	5%	3%
SRS	9	12%	4%	3%
Sonstige	34			
Total	76	100%	37%	23%
Methods				
Umfragen, Feedback-Bögen	35	20%	17%	11%
<i>Nicht spezifizierte Datenerhebungsmethoden</i>	32	18%	16%	10%
<i>Nennung spezifischer Indikatoren</i>	22	13%	11%	7%
IOOI	16	9%	8%	5%
Interviews	9	5%	4%	3%
Kooperationen mit Universitäten	10	6%	5%	3%
ToC	9	5%	4%	3%
externe Bewertungen bzw Evaluationen	7	4%	3%	2%
Beobachtungen	6	3%	3%	2%
Customer/user feedback	6	3%	3%	2%
Sonstige	21			
Total	173	100%	85%	53%
Certifications				
B-Corp	5	38%	2%	2%
Sonstige	8			
Total	13	100%	6%	4%

Wir haben die „Tools & Frameworks“ in einer Gruppe zusammengefasst, weil sie sich nur schwer voneinander trennen ließen. Auch ist die Zuordnung von „Nicht spezifizierten Datenerhebungsmethoden“, „Nennung spezifischer Indikatoren“ und GWÖ (z.T. auch Zertifikat) nicht unproblematisch. Die drei bzw. vier Oberkategorien sind somit nicht vollständig trennscharf.

Mit 23 Nennungen – das sind 30% der Nennungen in der Kategorie **Tools & Frameworks** und 11 % aller Befragten die angeben, ihre Wirkung zu messen bzw. 7 % aller DSEM-Befragten – wurde Phineo, das Kursbuch Wirkung bzw. die Wirkungstreppe am häufigsten genannt, gefolgt von der Gemeinwohl-Ökonomie bzw. Bilanz und SRS. Nur wenige Befragte gaben an, eigene Tools/Frameworks entwickelt zu haben. Auffällig ist der Fokus auf vor allem im deutschsprachigen Raum bekannte Tools und Frameworks (Phineo, SRS, GWÖ) während international geläufigere Tools und Frameworks selten genannt werden (z.B. IRIS+, GRI, IMP/Impact Frontiers). Gar nicht genannt wurden Tools zum Datenmanagement im engeren Sinne, also Software/SaaS, abgesehen von Tabellenkalkulationsprogrammen, obwohl zahlreiche auf IMM spezialisierte Angebote auf dem Markt vorhanden sind.

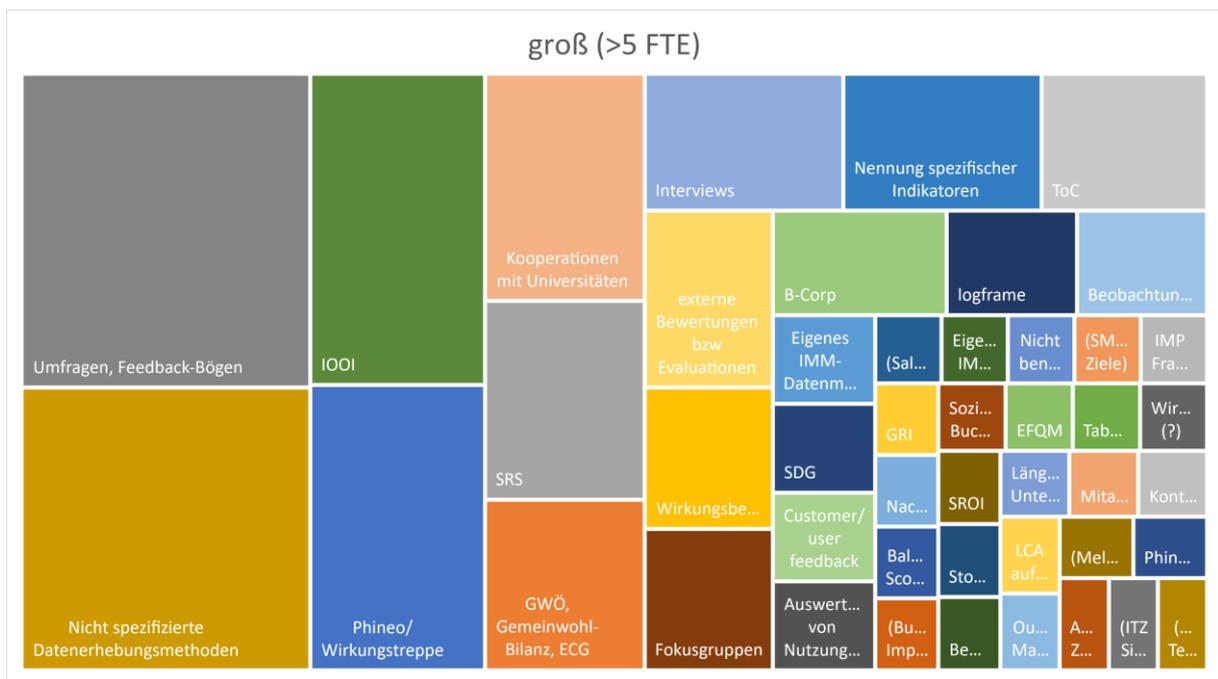
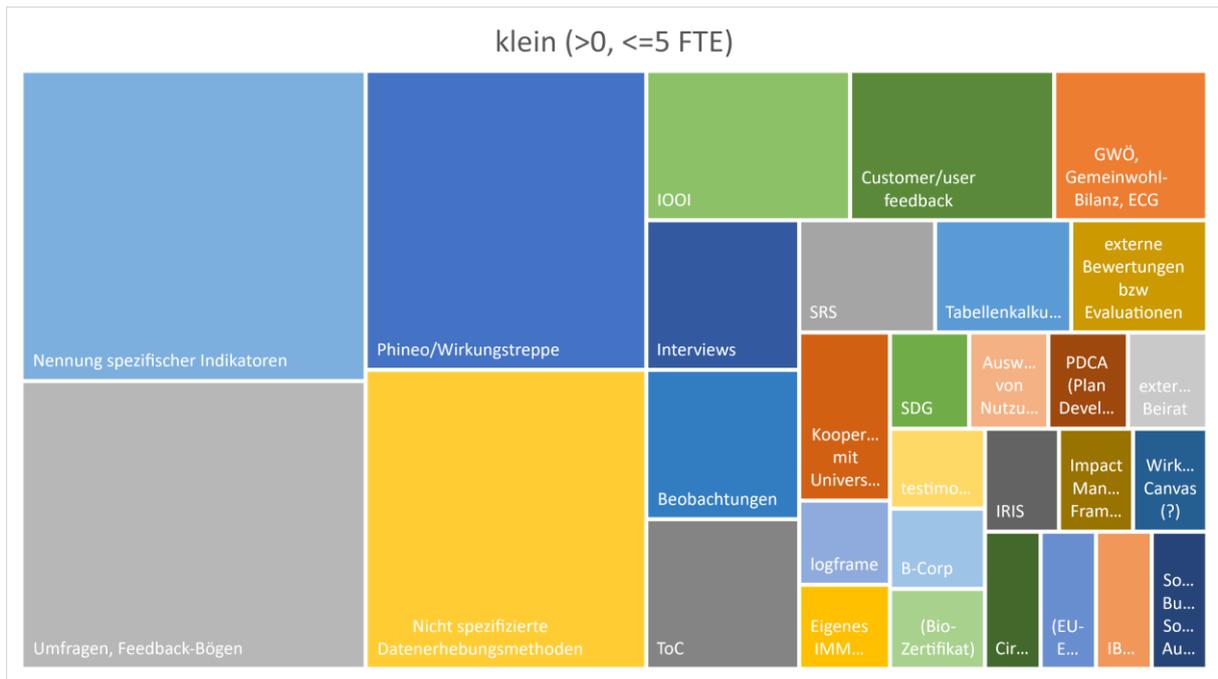
Bei den **Methoden** dominieren die Umfragen und Feedback-Bögen. Zudem wurden IOOI und Theory of Change (auch: logframe) relativ häufig genannt. Überraschend ist der relativ hohe Anteil der Nennungen (10) von Kooperationen mit Universitäten. Komplexere Wirkungsmessungsmethoden wie SROI, Outcome Mapping, Längsschnitt-Untersuchungen, Kontrollgruppendesign oder LCA (wohl: Life-Cycle-Analysis) wurden hingegen nur sehr vereinzelt genannt.

Zertifizierungen, obwohl in der Frage explizit erwähnt, spielen zahlenmäßig eine besonders untergeordnete Rolle. Die Zertifizierung als B-Corp nannten lediglich 5 Befragte, gefolgt von dem Phineo Wirkt-Siegel und AZAV mit je nur einer Nennung. Viele der sonstigen Zertifizierungen sind keine Wirkungs-Zertifizierungen im engeren Sinne. (Allerdings könnte auch die als *Framework* erfasste Angabe „GWÖ“ als Zertifizierung verstanden werden.)

Eine Visualisierung der Häufigkeit aller Nennungen von Tools, Frameworks, Methoden und Zertifizierungen unabhängig von der Oberkategorie bietet die Treemap im Anhang.

Unterschiede nach Größe der Organisation

Da Wirkungsmessung aufwendig ist, liegt die Annahme nahe, dass größere Organisationen andere Angaben zu den Tools, Frameworks, Methoden und Zertifizierungen gemacht haben, als kleine Organisationen. Die Umfrage beinhaltet Items zum Organisationsbudget, die jedoch von zahlreichen Befragten nicht ausgefüllt worden sind. Daher nutzen wir stattdessen die Angabe der Anzahl der Mitarbeiter (Q41 How many people are employed (and paid) by your organisation? - Number of FTE (full-time equivalent) positions (based on the standard working week in your country). Leider differenziert der Datensatz nicht zwischen *missing* und der Angabe „Null“ (beides als „0“ codiert). Wir lassen diese Fälle daher aus (und somit 47 Organisationen die entweder keine bezahlten Mitarbeiter:innen haben, oder diese Zahl nicht angeben wollten). Die übrigen Teilen wir in zwei Gruppen, ein: Bis zu 5 FTE in „klein“ (146 Fälle) und 5-3000 FTE in „groß“ (156 Fälle).



Kleinere Organisationen listen häufiger spezifische Output- oder Outcome-Indikatoren auf und erwähnten seltener Kooperationen mit Universitäten sowie externe Evaluationen, was nicht überrascht. Andererseits machten große Organisationen relativ häufig unspezifische/vage Angaben zu Datenerhebungsmethoden. SRS und GWÖ finden sich stärker bei größeren Organisationen wieder, wurden aber auch von kleineren genannt. Phineo/Wirkungstreppe hat mit jeweils 11 in beiden Gruppen genau gleich viele Nennungen. Bei den Zertifikaten gibt nur eine kleinere Organisation die B-Corp Zertifizierung an gegenüber 4 großen. Auch sind IOOI und ToC bei größeren Organisationen scheinbar bereits stärker etabliert als bei kleineren.

Zusammenfassend lässt sich zur Frage des Unterschieds zwischen kleinen und großen Organisationen sagen, dass sich am *Ranking* der Nennungen von Tools & Frameworks, Methoden und Zertifizierungen innerhalb dieser drei Oberkategorien nur geringe Unterschiede ausmachen lassen

(siehe Tabelle 3). Allerdings weisen die kleineren Organisationen mit 0,6 Nennungen pro Befragten einen deutlich niedrigeren Wert auf, als große Organisationen, die durchschnittlich 1,2 Tools & Frameworks, Methoden und Zertifizierungen benannt haben: Große Organisationen haben im Durchschnitt doppelt so viele Angaben zu Wirkungsmessungsmethoden, -tools, -frameworks und -zertifizierungen gemacht, wie kleine Organisationen, obgleich es sich hierbei nicht immer um Wirkungsmessung im Sinne der Evaluationsforschung und -praxis handelt.

Schlussfolgerungen

Der DSEM 2023-2024 bietet wertvolle Erkenntnisse über die Verbreitung von Praktiken der Wirkungsmessung und des Wirkungsmanagements (IMM) in der *social economy* Deutschlands, denn systematisch erhobene und belastbare Daten zu dieser Frage sind bisher rar.¹ Sie lassen die vorläufige Schlussfolgerung zu, dass zwar ein großer Teil der Unternehmen und Organisationen (noch) keine explizite Wirkungsmessung betreibt, sich aber zugleich ein beachtliches Feld an „Pionieren“ ausmachen lässt, die wiederum sehr unterschiedlich auf die im DSEM gestellte Frage nach den „*tools, methods, frameworks, certifications or approaches*“ geantwortet haben: Während ein Teil der Befragten konkrete individuelle Indikatoren und Methoden auflistet, verweisen andere auf vorhandene Frameworks. In Bezug auf letztere hat sich offenbar die Theory of Change sowie die IOOI-Logik durchgesetzt, insbesondere in derjenigen Form, die Phineo bzw. die Bertelsmann Stiftung mit der Wirkungstreppe und den Social Reporting Standards im deutschsprachigen Raum erfolgreich eingeführt haben. Eine ebenfalls relevante Menge setzt hingegen auf das ebenfalls im deutschsprachigen Raum v.a. bei mittelgroßen gemeinwohlorientierten Unternehmen bekannte Programm der Gemeinwohlökonomie, die sich von der Wirkungstreppe in vielerlei Hinsicht unterscheidet – unter anderem darin, dass sie einen normativen Rahmen zur Bewertung der Wirkung (z.B. Menschenrechte) bietet bzw. vorgibt.

Auffällig ist zudem, was in den Antworten nicht genannt wird, wenn man diese mit dem breiten internationalen Feld der Wirkungsmessung und -evaluation in Wissenschaft und Praxis vergleicht: Social Return on Investment (SROI) als relativ umfassende Methode der Wirkungsmessung und -monetarisierung wurde lediglich ein einziges Mal genannt. Internationale Alternativen zur Wirkungstreppe als Konkretisierung von ToC/IOOI wie z.B. die „five dimensions of impact“ des Impact Management Project (bzw. Impact Frontiers, übernommen z.B. in IRIS+) bleiben ebenfalls außen vor. Abgesehen von B-Corp spielen Zertifikate eine untergeordnete Rolle. Der gesamte Bereich der Impact Valuation, also der Umrechnung von Wirkungen in Geldeinheiten, taucht ebenfalls nicht auf. Interessant ist die Nennung von Kooperationen mit Universitäten von immerhin 10 Befragten: Dies könnte auf eine wichtige Bedeutung des Transfers aus der Forschung in die Praxis hinweisen.

Die augenscheinliche Dominanz der Wirkungstreppe hat potenziell inhaltliche Auswirkungen auf die Wirkungsmessungspraxis in Deutschland: Anders als z.B. bei Impact Frontiers oder SROI spielt die Frage der Attribution, also der kausalen Zurechnung von beobachteten Veränderungen zur Aktivität der Organisation anhand des Vergleichs mit dem Counterfactual (*sine qua non*), bei der

¹ Zielinski, F., Mildenerger, G., Rabadjieva, M., Sauerbier, E., Terstriep, J., & Wruk, D. (2024). Was sind Soziale Innovationen und wie lassen sich ihre Impacts messen? Konzeptionelle Grundlagen für innovationsfeldspezifische Wirkmodelle. Research Paper, Centrum für Soziale Investitionen und Innovationen, Universität Heidelberg, S. 32-34. doi.org/10.11588/heidok.00035636

Wirkungstreppe keine wesentliche Rolle. Damit verliert die Unterscheidung zwischen Outcomes und Impacts an Trennschärfe, was bei der Gestaltung von organisationalen Wirkmodellen in der Praxis zu Problemen führen dürfte. Andererseits umgeht man auf diese Weise die Herausforderungen in Bezug auf Aufwand, Know-how und Datenverfügbarkeit, die sich mit dem Counterfactual-basierten Verständnis der Unterscheidung Impact/Wirkung ergeben.²

² Für einen Vorschlag eines graduellen Wirkmodells das beide Verständnisse von Impact integriert siehe: Ebd. S. 39ff.

Anhang

Tabelle 2

Gesamtauswertung	Mehrfachnennungen möglich			
	Anzahl Nennungen	Anteil Nennungen	Anteil an Respondenten mit Wirkungsmessung	Anteil an Respondenten insgesamt
Tools & Frameworks				
Phineo/Wirkungstreppe	23	30%	11%	7%
GWÖ, Gemeinwohl-Bilanz, ECG	10	13%	5%	3%
SRS	9	12%	4%	3%
Wirkungsbericht	4	5%	2%	1%
Eigenes IMM-Datenmanagement-Tool	3	4%	1%	1%
Tabellenkalkulation	3	4%	1%	1%
SDG	3	4%	1%	1%
Soziale Buchführung / Sozial-Audit	2	3%	1%	1%
Wirkungsspinne (?)	1	1%	0%	0%
PDCA (Plan Develop Check Act?)	1	1%	0%	0%
(Meltwater Media Intelligence)	1	1%	0%	0%
(Salesforce)	1	1%	0%	0%
Eigenes IMM System entwickelt	1	1%	0%	0%
Nicht benanntes IMM-Datenmanagement-Tool	1	1%	0%	0%
IRIS	1	1%	0%	0%
Impact Management Framework von Impact Hub (?)	1	1%	0%	0%
(SMART-Ziele)	1	1%	0%	0%
Wirkungsmodell-Canvas (?)	1	1%	0%	0%
IMP Framework	1	1%	0%	0%
CirculaRise	1	1%	0%	0%
GRI	1	1%	0%	0%
Nachhaltigkeitsberichterstattung	1	1%	0%	0%
EFQM	1	1%	0%	0%
Balanced Scorecard BSC	1	1%	0%	0%
(Budget Impact Modeling)	1	1%	0%	0%
(EU-Emissionshandel)	1	1%	0%	0%
IBMS	1	1%	0%	0%
Total	76	100%	37%	23%
Methods				
Umfragen, Feedback-Bögen	35	20%	17%	11%
<i>Nicht spezifizierte Datenerhebungsmethoden</i>	32	18%	16%	10%
<i>Nennung spezifischer Indikatoren</i>	22	13%	11%	7%
IOOI	16	9%	8%	5%
Interviews	9	5%	4%	3%
Kooperationen mit Universitäten	10	6%	5%	3%
ToC	9	5%	4%	3%
externe Bewertungen bzw Evaluationen	7	4%	3%	2%
Beobachtungen	6	3%	3%	2%
Customer/user feedback	6	3%	3%	2%
logframe	4	2%	2%	1%
Fokusgruppen	4	2%	2%	1%
Auswertung von Nutzungsdaten/interner Prozessdaten	3	2%	1%	1%
SROI	1	1%	0%	0%
Storytelling	1	1%	0%	0%
Benchmarking	1	1%	0%	0%
Längsschnitt-Untersuchungen	1	1%	0%	0%
Mitarbeitergespräche	1	1%	0%	0%
Kontrollgruppendesign	1	1%	0%	0%
externer Beirat	1	1%	0%	0%
LCA auf Produktebene	1	1%	0%	0%
Outcome Mapping	1	1%	0%	0%
testimonials	1	1%	0%	0%
Total	173	100%	85%	53%
Certifications				
B-Corp	5	38%	2%	2%
Phineo Wirkt Siegel	1	8%	0%	0%
AZAV-Zertifizierung	1	8%	0%	0%
(FairWear)	1	8%	0%	0%
(ITZ Siegel)	1	8%	0%	0%
(Enactus Environment)	1	8%	0%	0%
(Öko-Tex)	1	8%	0%	0%
(Bio-Zertifikat)	1	8%	0%	0%
(GOTS)	1	8%	0%	0%
Total	13	100%	6%	4%

Tabelle 3

Gesamtauswertung - nach Größe		
	Mehrfachnennungen möglich	
	Anzahl Nennungen	
Tools & Frameworks	klein (>0, <=5 FTE)	groß (>5 FTE)
Phineo/Wirkungstreppe	11	11
GWÖ, Gemeinwohl-Bilanz, ECG	3	6
SRS	2	7
Wirkungsbericht	0	4
Eigenes IMM-Datenmanagement-Tool	1	2
Tabellenkalkulation	2	1
SDG	1	2
Soziale Buchführung / Sozial-Audit	1	1
Wirkungsspinne (?)	0	1
PDCA (Plan Develop Check Act?)	1	0
(Meltwater Media Intelligence)	0	1
(Salesforce)	0	1
Eigenes IMM System entwickelt	0	1
Nicht benanntes IMM-Datenmanagement-Tool	0	1
IRIS	1	0
Impact Management Framework von Impact Hub (?)	1	0
(SMART-Ziele)	0	1
Wirkungsmodell-Canvas (?)	1	0
IMP Framework	0	1
CirculaRise	1	0
GRI	0	1
Nachhaltigkeitsberichterstattung	0	1
EFQM	0	1
Balanced Scorecard BSC	0	1
(Budget Impact Modeling)	0	1
(EU-Emissionshandel)	1	0
IBMS	1	0
Total	28	46
Methods	klein	groß
Umfragen, Feedback-Bögen	13	20
Nicht spezifizierte Datenerhebungsmethoden	11	18
Nennung spezifischer Indikatoren	14	6
IOOI	4	12
Interviews	3	6
Kooperationen mit Universitäten	2	8
ToC	3	5
externe Bewertungen bzw Evaluationen	2	5
Beobachtungen	3	3
Customer/user feedback	4	2
logframe	1	3
Fokusgruppen	0	4
Auswertung von Nutzungsdaten/interner Prozessdaten	1	2
SROI	0	1
Storytelling	0	1
Benchmarking	0	1
Längsschnitt-Untersuchungen	0	1
Mitarbeitergespräche	0	1
Kontrollgruppendesign	0	1
externer Beirat	1	0
LCA auf Produktebene	0	1
Outcome Mapping	0	1
testimonials	1	0
Total	63	102
Certifications	klein	groß
B-Corp	1	4
Phineo Wirkt Siegel	0	1
AZAV-Zertifizierung	0	1
(FairWear)	0	0
(ITZ Siegel)	0	1
(Enactus Environment)	0	0
(Öko-Tex)	0	1
(Bio-Zertifikat)	1	0
(GOTS)	0	0
Total	2	8

