

Deutsche Gesellschaft für Evaluation (DeGEval 2004)

**„Befragung zur Steuerung des Revisionsprozesses der
DeGEval-Standards für Evaluation (REVAL)
- Bericht zur qualitativen Datenauswertung -**

Koordination für den DeGEval-Vorstand und wissenschaftliche Begleitung
Wolfgang Beywl

Durchführung und Berichterstattung
Henry Kotitschke

Inhalt:

Einführung / Erläuterungen zur Auswertung	2
Interpretationsgrundlage:	2
Das Kategoriensystem	3
Besonderheiten bei der Auswertung der qualitativen Daten bei REVAL.....	4
Erläuterungen zu Kategorie III (Kommentare / Anmerkungen zu den Standards und dem Revisionsprozess):	5
Auswertung der qualitativen Daten	5
Kategorie I: Nutzung der Standards	5
Kategorie II: Begründungen für Revisionsalternativen	9
Kategorie III: Kommentare / Anmerkungen zu den Standards und dem Revisionsprozess	17
Issues	35
Ergebniszusammenfassung	37

Einführung / Erläuterungen zur Auswertung

Die „Befragung zur Steuerung des Revisionsprozesses der DeGEval-Standards für Evaluation“ wurde als Online-Befragung im Zeitraum vom 02.04.2004 bis zum 02.06.2004 durchgeführt. Die Umfrage soll zum einen die Meinung und fachliche Position der DeGEval-Mitglieder zur Thematik „Revision der Standards“ erfassen, zum anderen auch die Haltung sonstiger Interessierter. Im Besonderen spricht die Befragung Personen an, die Evaluationen in Auftrag geben oder Evaluationen durchführen.

Die Ergebnisse der Befragung sollen eine Entscheidungsgrundlage über das Ob, Wann und Wie einer Revision der Standards darstellen.¹

Im Rahmen der quantitativen Analyse konnte bereits ein Ergebnis formuliert werden, dass Aufschluss hinsichtlich der Dringlichkeit einer Revision und der Bereitschaft der Befragungsteilnehmenden zur Mitarbeit an einem Revisionsprozess gibt.²

Die vorliegende qualitative Datenauswertung kann nunmehr in Ergänzung zu den quantitativen Ergebnissen weitere tiefergehende Erkenntnisse zum Thema Revision der Standards liefern. In diesem Sinne ist es beispielsweise möglich, zu analysieren, mit welcher Begründung sich die Teilnehmenden für oder gegen eine der im Online-Fragebogen vorgeschlagenen Revisionsalternativen entscheiden. Weiterhin können Erkenntnisse zur Nutzung der DeGEval-Standards und eine große Vielfalt an Verbesserungsvorschlägen gewonnen werden. Auf Grundlage dieses Wissens kann die inhaltliche Ausrichtung einer Revision zweckgerichtet vorgenommen werden.

Interpretationsgrundlage:

Die gesammelten qualitativen Daten der „Befragung zur Steuerung des Revisionsprozesses der DeGEval Standards für Evaluation“ (REVAL) werden im Folgenden auf Grundlage der inhaltlichen Akzentuierung der entsprechenden Fragen und den Erkenntnissen aus der quantitativen Auswertung analysiert und interpretiert.

Der Analyse zugrunde liegt die Antwortgesamtheit von vier offenen Fragen:

1. 4. In welcher Hinsicht sind die Standards für Ihre Arbeit nützlich?
(bitte skizzieren Sie kurz, für welche Zwecke Sie die Standards nutzen)

.....
217 Antwortbeiträge

2. *Bitte begründen Sie kurz Ihre Aussage:*.....

Gemeint ist das Antwortverhalten bei Frage 6:

„Welche der vorgeschlagenen Alternativen für eine Revision bevorzugen Sie? (Der Revisions- bzw. Entwicklungsaufwand steigt von 1 zu 4 erheblich an.) Plädieren Sie für eine weitere Option?“

(1) Eine Revision sollte sich ausschließlich auf den DeGEval-Standard-Text (die 25 Einzelstandards) beziehen.

¹ Für weitere Informationen über die Vorgeschichte und den Zweck der Evaluation, sowie zur technischen Umsetzung und Durchführung der Befragung greifen Sie bitte auf den Bericht zur quantitativen Datenauswertung von REVAL zurück.

² Eine deutliche Mehrheit der Teilnehmenden spricht sich *für* eine Revision aus (130 von 177 Personen, die eine Position zur Revision einnehmen.). Dabei wird eine mittelfristig anzustrebende Revision (bis 2008) leicht gegenüber einer kurzfristigen (bis 2005) vorgezogen. Wenn es zu einer Revision kommt, ist eine deutliche Mehrheit für eine Revision der gesamten Broschüre bzw. für eine bezüglich Inhalt und Umfang deutlich erweiterte Fassung. Weiterhin besteht eine große Bereitschaft, an der Revision der Standards mitzuwirken. Nur 26,4% der Teilnehmenden, die eine Revision befürworten, versagen ihre Mitarbeit.

- (2) Eine Revision sollte sich ausschließlich auf die Erläuterungen beziehen
- (3) Eine Revision sollte sich auf den gesamten Inhalt der Broschüre beziehen
- (4) Eine Revision sollte in Inhalt und Umfang deutlich über die Broschüre „Standards für Evaluationen“ hinausgehen (z.B. ein Produkt anstreben, das der Buchfassung der Standards des „Joint Committee“ mit zusätzlichen „Richtlinien“, „Häufigen Fehlern“ sowie kommentierten Anschauungsbeispielen entspricht).
- (5) Eine Revision der Broschüre halte ich z.Z. für nicht erforderlich
- (6) sonstiges:.....
- (7) weiß nicht, bin unsicher

152 Antwortbeiträge

3. 8. Was könnte Ihrer Auffassung nach verbessert werden? Welche Veränderung/Erweiterung/Verbesserung schlagen Sie vor? *(erläutern Sie kurz ihre Vorstellungen)*

.....
95 Antwortbeiträge

4. 15. Evtl. Anmerkungen/Kommentare zum Thema Standards/Revisionsprozess

.....
44 Antwortbeiträge

Die Einträge frei ausfüllbarer Antwortmöglichkeiten („sonstiges“ Eingabefelder) werden hier nicht berücksichtigt. Deren Betrachtung ist bereits im Rahmen der quantitativen Auswertung vorgenommen worden.

Insgesamt haben 212 der 257 Befragten (82,5%) die Gelegenheit wahrgenommen, auf mindestens eine der offenen Fragen zu antworten. Alles in allem liegen 508 Beiträge vor, die zusammen 106 874 Zeichen enthalten (ohne Leerzeichen).

Die Auswertung der qualitativen Daten wurde mithilfe der Text-Sortier-Technik (TST) durchgeführt.³ Dabei wurden die Antwortbeiträge der Befragten in z.T. mehrere Sinneinheiten aufgegliedert und in einem für diesen Zweck erstellten Kategoriensystem abgelegt. Die dem Kategoriensystem immanente logische Struktur wurde in Bezug zu den entstandenen Sinneinheiten und der Eigenart der jeweiligen Frage entwickelt.

Sämtliche Bemühungen der Analyse und Interpretation, gehen auf die Datengesamtheit des entwickelten Kategoriensystems zurück.

Das Kategoriensystem

Die Antwortgesamtheit des Kategoriensystems ist in drei große Kategorien aufgeteilt worden. Der unterschiedlichen Ausrichtung der Fragen entsprechend, existiert eine Kategorie mit Sinneinheiten zum Thema „Nutzung der Standards“, eine zum Thema „Begründungen für bestimmte Revisionsalternativen“, sowie eine zum Thema „Kommentare / Anmerkungen zu den Standards und dem Revisionsprozess“. Diese grundlegende Struktur ermöglicht es nahezu alle Antwortbeiträge aus den oben genannten offenen Fragen entsprechend zu kategorisieren.

³ Die TST ist eine Auswertungstechnik für qualitative Daten auf Basis eines konventionellen Textverarbeitungsprogramms; sie wurde Ende der 80er Jahre von Wolfgang Beywl entwickelt.

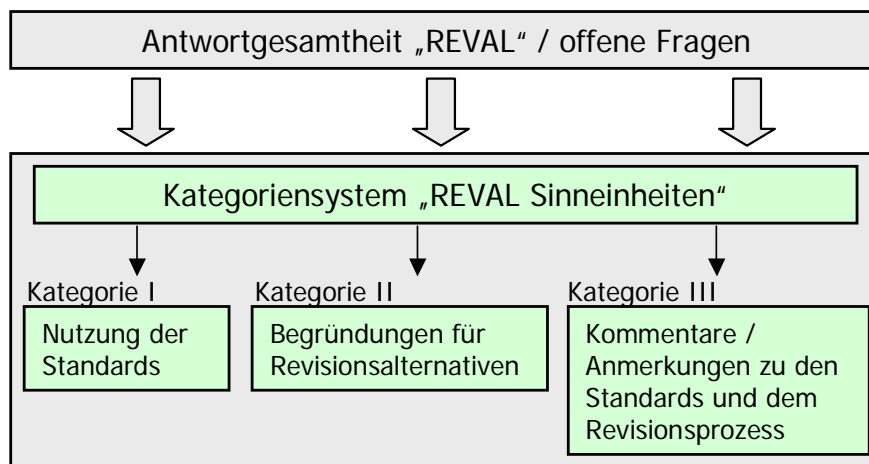


Abbildung 1: Qualitative Auswertung von REVAL mithilfe eines Kategoriensystems

Die Analyse und Auswertung der qualitativen Daten wird sich im Folgenden ganz konkret an dieser Kategorienstruktur orientieren und infolgedessen in drei Abschnitte unterteilt.

Besonderheiten bei der Auswertung der qualitativen Daten bei REVAL

Das Antwortverhalten der Befragten weist eine Struktur auf, die für die Auswertung der qualitativen Daten berücksichtigt werden sollte.

Abgesehen von den Antworten bzw. Sinneinheiten der Kategorie I (Nutzung der Standards), welche sich explizit auf die jeweilige Verwendung und den Nutzen der DeGEval-Standards beziehen, stellen sich die restlichen Antwortbeiträge der Befragung weniger spezifisch dar. Eine Differenzierung der Antworten hinsichtlich der inhaltlichen Ausrichtung der Fragen (intendierter Fragenzweck) ist dadurch erschwert worden:

Die freien Textfelder bei den Fragen 6, 8 und 15 wurden in vielen Fällen unabhängig vom Sinn und Zweck der jeweils gestellten Frage verwendet. Oder anders formuliert: Die Fragen sind in ihrer Ausrichtung nicht trennscharf genug verfasst worden. Diesem Umstand zufolge können viele der Antwortbeiträge gleich mehreren Fragen, und infolgedessen auch Kategorien, zugeordnet werden.

Bei der Konstruktion des Kategoriensystems ist diese Besonderheit folgendermaßen berücksichtigt worden:

1. Die Antworten auf die offenen Fragen 8 und 15 werden in einer Kategorie zusammengefasst und jeweils dort differenziert gegliedert. Die Aufteilung der Antworten dieser beiden Fragen in zwei entsprechende, separate Kategorien macht aufgrund der Ähnlichkeit der Antworten keinen Sinn.
Markante Gliederungspunkte: Verbesserungsvorschläge (allgemein / konkret), Kommentare / Bewertungen zu den Standards / dem Revisionsprozess
2. Die Antworten auf die offene Frage 4 werden in einer gesonderten Kategorie „Nutzung der Standards“ gesammelt und gegliedert. Dies empfiehlt sich, da diese Frage ein sehr spezifisches Antwortpotenzial enthält.
3. Die Beiträge aus dem Textfeld bei Frage 6 werden in einer gesonderten Kategorie „Begründungen für Revisionsalternativen“ gesammelt. Diese Kategorie macht deshalb Sinn, da die jeweiligen Begründungen in Zusammenhang mit der jeweils begründeten Revisionsalternative interpretiert werden sollten. Diese Zuordnung von Begründung zur entsprechenden Alternative ist nur im Zusammenspiel mit den quantitativen Daten von Frage 6 möglich. Bei den offenen Antworten der

Fragen 4, 8 und 15 ist eine Herleitung zur Entscheidung für eine bestimmte Revisionsalternative nur mit großem Arbeitsaufwand möglich und wird nicht betrieben.

Neben den Kategorien I und II, die eine spezifische Sammlung von Antwortbeiträgen hinsichtlich der Fragenvorgaben beinhalten, existiert mit Kategorie III eine sehr umfangreiche Kategorie, die zunächst einiger weiterer Erläuterungen bedarf.

Erläuterungen zu Kategorie III (Kommentare / Anmerkungen zu den Standards und dem Revisionsprozess):

Die Fragen nach Verbesserungs-, Erweiterungs- und Veränderungsvorschlägen (Frage 8) und Anmerkungen zu den Standards bzw. dem Revisionsprozess (Frage 15) verursachen relativ ähnliche Antwortprofile. Aufgrund dieser Ähnlichkeit wurde eine gemeinsame Antwortkategorie entwickelt, die in Form fein ausdifferenzierter Unterkategorien einen vielschichtigen Blick auf die im Zentrum dieser Fragen stehenden Schwerpunkte ermöglicht.

Sowohl Verbesserungsvorschläge, als auch Anmerkungen und Kommentare implizieren eine Auseinandersetzung mit dem Gegenstand / Sachverhalt. Im Falle eines Verbesserungsvorschlags wird dann noch einen Schritt weitergegangen, und aus dieser Auseinandersetzung eine Konsequenz abgeleitet, welche in einen konkreten Vorschlag mündet. Der Grad der Konkretheit variiert stark, sodass einige Verbesserungsvorschläge in ihrer Aussagekraft nicht viel mehr vermitteln, als ein gut formulierter Kommentar. Dieser Systematik folgend, kann man die Bildung einer vereinigenden Kategorie auch so verstehen, dass die Kommentare in einer logischen Konsequenz durch die Verbesserungsvorschläge ergänzt werden.

In diesem Sinne ist ein Gliederungssystem für die Doppelkategorie „Kommentare / Anmerkungen zu den Standards und dem Revisionsprozess“ entwickelt worden, indem Verbesserungsvorschläge und Kommentare gleichermaßen aufgenommen sind.

Die Verhältnismäßigkeiten der Antworten in der Kategorie III entsprechen nicht zwangsläufig den Verhältnissen der quantitativen Analyse. Es handelt sich sozusagen um ein gesondertes Mehrheitsverhältnis von Beiträgen zum Thema „Revision der DeGEval-Standards“, die unabhängig von der Entscheidung für oder gegen eine bestimmte Revisionsalternative interpretiert werden.

Auswertung der qualitativen Daten

Die Analyse der qualitativen Daten kann in Ergänzung zu den quantitativen Daten auf verschiedene Schwerpunkte der Befragung einen differenzierten Blick ermöglichen.

Kategorie I: Nutzung der Standards

KATEGORIE	NENNUNGEN
100 – allgemeine Orientierungsfunktion der Standards	35
101 – Standards als Anregung und Ausgangspunkt (Weiterentwicklung)	16
110 – Standards als phasenübergreifendes Hilfsmittel bei Evaluationsprojekten	46
111 - Standards als Hilfsmittel zur Konzeption von Evaluationen / Auftragsklärung	53
112 – Standards als Hilfsmittel bei der Durchführung von Evaluationen	32
113 – Standards als Hilfsmittel bei der Auswertung und Berichterlegung von Evaluationen	5
120 – Standards als Element der Qualitätskontrolle / Standards als Grundlage zur Bewertung von Evaluationen	51
121 – Standards als Hilfsmittel zur Selbstreflexion	13
122 – Standards als Grundlage für Metaevaluationen	20

123 – Standards als Referenz / Grundlage bei wissenschaftlichen Arbeiten	17
125 – Verwendung der Standards innerhalb der Beratung von Evaluationsprojekten	6
130 – Standards als Argumentationsgrundlage / Standards als Mittel zur Legitimation	41
131 – Standards als Grundlage zur Schaffung einer vertrauensbildenden Basis	12
132 – State of Art – Standards als Ausweis für Professionalität	25
133 – Verwendung der Standards zu Werbezwecken / Marketing	6
134 – explizite Verwendung der Standards gegenüber Auftraggebenden	25
135 – Standards als ethisches Schutzschild	3
140 – Standards als Lehrgrundlage	42
141 – Verwendung der Standards zur Einarbeitung neuer Mitarbeiter / Standards als Hilfe für Evaluationsneulinge	10
142 – Standards als Lehrgrundlage zum Umgang mit Beteiligten und Betroffenen (teach the stakeholders)	5
143 – Standards als Lehrgrundlage zum Umgang mit Auftraggebenden (teach the client)	4
150 – spezifische Einsatzfelder der Standards	18
151 – Verwendung der Standards zu Zwecken der Selbstevaluation	6
152 – Verwendung der Standards im Bereich Bildung	11
160 – benachbarte Standardsets	2
161 – noch keine Nutzung	12

Aus den 217 Antworten auf die Frage 4 des Fragebogens (In welcher Hinsicht sind die Standards für ihre Arbeit nützlich?) sind 516 Sinneinheiten abgeleitet worden, welche einen differenzierten Blick auf das Thema „Nutzung der Standards“ ermöglichen.

Zunächst soll die Bedeutung der Standards als **allgemeine Orientierungsgrundlage** hervorgehoben werden. 35 Nennungen drücken klar aus, dass die Standards eine grundlegende Orientierung im Bereich Evaluation gewährleisten. Sie stellen einen generellen Rahmen dar, in welchem sich Evaluationen abspielen („Gedankenraster“), sie erleichtern die Kommunikation in diesem Bereich („Orientierung in der Evaluationsdebatte“), sie verschaffen einen Überblick („Um sich einen standardisierten Überblick zu verschaffen“) und dienen als Nachschlagewerk: „Jedes Mal, wenn ich bei meiner Tätigkeit (Auftraggeber, Koordination) mit Fragen der Evaluation in Kontakt komme, dienen die Standards mit Erläuterungen mir als Orientierung und Nachschlagewerk.“. Zudem fällt, dank der Standards, die begriffliche Orientierung in Sachen Evaluation wesentlich leichter („begriffliche Orientierung“). Neben den Begrifflichkeiten erleichtern die Standards auch die Definition möglicher Ziele und Zwecke von Evaluationen („Definition der Zwecke/Funktionen von Evaluation etc.“), sodass der Aussage einer / eines Teilnehmenden im Sinne eines zusammenfassendes Statements, weitgehend zugestimmt werden kann: „Die Standards sind Arbeitsgrundlage und 'Messlatte' für jegliche Befassung mit Belangen der Evaluation.“

In dieser grundlegenden Funktion stellen die Standards sowohl eine Grundlage / einen Ausgangspunkt für die Weiterentwicklung der Standards selbst, als auch für andere Bemühungen in Wissenschaft und Forschung dar (siehe Kategorie 101, 16 Nennungen).

Der Aufgabe der allgemeinen Orientierung wird durch die **Verwendung der Standards als Phasen übergreifendes Hilfsmittel in konkreten Evaluationsprojekten**, ein weiterer Aspekt der Standardsnutzung hinzugefügt. Die Nutzung der Standards in diesem Sinne, stellt wesentlich konkretere Anforderungen an die Standards, als die bloße Orientierungsfunktion. Gleichwohl ist diese Art der Nutzung, eine der am häufigsten genannten: Metakategorie 110 (Standards als phasenübergreifendes Hilfsmittel bei Evaluationsprojekten) vereint, zusammen mit seinen Untergliederungen 111 – 113, immerhin 136 Nennungen. Die

Untergliederungen spiegeln die unterschiedlichen Phasen eines Evaluationsprojekts wieder, in welcher die Standards als Hilfsmittel gefordert werden. Selbstverständlich sind die Standards im besten Fall in allen drei Phasen einer Evaluation (1. Konzeption / Planung → 2. Durchführung → 3. Berichtlegung) im Sinne eines Hilfsmittels wirksam, so wie dies auch von einigen Befragten beschrieben wird: **„Als roter Faden während der Kontraktverhandlungen, der Erhebung, der Datenausarbeitung und der Rückmeldung.“** Allerdings liegen die meisten Nennungen im Bereich „Konzeption von Evaluationen / Auftragsklärung“ (Kategorie 111) vor. 53 Beiträge in dieser Kategorie deuten erneut die theoretische Orientierungsfunktion der Standards an. Im Fall der Beiträge von Kategorie 111 werden jedoch ganz konkrete Handlungsanweisungen aus den Standards abgeleitet. Neben der ersten Phase einer Evaluation, werden die Standards von den Befragten auch bei der Durchführung von Evaluationen (Kategorie 112, 32 Nennungen) sowie bei der Berichtlegung (Kategorie 113; 5 Nennungen) ebensolcher verwendet. In vielen weiteren Beiträgen wird die Nutzung der **Standards als Werkzeug zur Qualitätssicherung** (Metakategorie 120, 51 Nennungen) ausgeweitet.

„Ich arbeite als Organisationsentwicklerin häufig mit Verfahren der Qualitätsentwicklung/Qualitätsmanagement, vor allem an EFQM - angelehnte Formen der organisationsinternen Selbstbewertung. Ich empfehle meinen Kunden, sich für die Planung der Selbstbewertung an den Standards zu orientieren und nutze sie natürlich auch selbst.“

Kontrolle und Bewertung von Evaluationsbemühungen in den verschiedensten Bereichen, in denen Evaluation zum Einsatz kommt, ist einer der wichtigsten Verwendungszwecke für die Standards. Dabei werden die Standards sowohl zur nachträglichen Bewertung von Evaluationsstudien (Standards als Grundlage für Metaevaluationen; Kategorie 122, 20 Nennungen), als auch zur permanenten Kontrolle in laufenden Projekten genutzt. Abgesehen von der übergeordneten Bewertungsfunktion kompletter Projekte, werden die Standards auch punktuell zu verschiedenen Einzelzwecken der Qualitätssicherung angewendet: Beispielsweise liegen 17 Antwortbeiträge vor, in denen die Standards als Referenz / Grundlage bei wissenschaftlichen Arbeiten Einsatz finden (Kategorie 123). Außerdem dienen die Standards vielen Befragten als Hilfsmittel zur Selbstreflexion (Kategorie 121, 13 Nennungen). Schließlich finden die Standards innerhalb der Beratung von Evaluationsprojekten Anwendung (Kategorie 125, 6 Nennungen).

Eine weitere Verwendung der Standards, die von einer Vielzahl der Befragten angegeben wurde, ist die, der **Standards als Argumentationsgrundlage** bzw. die, der **Standards als Mittel zur Legitimation**. Die Kategorie 130, in der diese Beiträge aufgelistet sind, umfasst neben der Hauptkategorie, 5 Untergliederungen die diese Thematik teilen. Die Bedeutung der Legitimierungsfunktion ist nicht zu unterschätzen: Die Standards bürgen für standardisierte, nachprüfbare Vorgehensweisen. Bezieht man sich auf sie, stärkt das auch das eigene Vorgehen und die Position auf einem entsprechenden Markt: *Die Standards sind für meine Arbeit nützlich ...* „zur Legitimation von Qualitätsansprüchen an Evaluationsentwürfe und Verfahren (Beispiele: Beteiligung, Transparenz der Zwecke.“

Die gesammelten Beiträge in diesem Zusammenhang sind besonders deswegen interessant, weil sie nachweisen, dass die DeGEval-Standards ein anerkanntes Mittel zur Legitimation empirischer Vorgehensweisen darstellen, oder zumindest eine Argumentationsgrundlage aufzeigen, um eine Position / Vorgehensweise zu begründen. Die fünf Untergliederungen positionieren sich rings um diese Problematik: Die Legitimation von Vorgehensweisen im Rahmen von Evaluationsprojekten muss in den meisten Fällen gegenüber den Auftraggebenden

geltend gemacht werden. In diesem Zusammenhang gibt es viele Beiträge, die explizit die Verwendung der Standards gegenüber Auftraggebenden thematisieren („Gesprächsleitfaden mit dem Auftraggeber“; siehe Kategorie 134, 25 Nennungen). In einer ähnlichen Funktion werden die Standards von vielen Befragten als Grundlage zur Schaffung einer vertrauensbildenden Basis angesehen und verwendet (Kategorie 131, 12 Nennungen):

„Auch widerständige Beteiligte und Betroffen erhalten durch die Standards einen transparenten und kompakten Einblick in meinen Evaluationsansatz, was hilft, Vorurteile auszuräumen, insbesondere das Vorurteil, dass Evaluation nur zusätzliche Arbeit darstelle, notwendig für die Außendarstellung, aber nicht hilfreich für die Qualitätsentwicklung vor Ort.“

Hierbei ist nicht nur die Kommunikation zwischen Evaluierenden und Auftraggebenden eines Evaluationsprojektes gemeint, sondern sämtliche Interaktionen innerhalb verschiedenster Projektformate: „Zur Rechtfertigung (bei Anfeindungen) und Argumentation (bei Konzeptionen)“

Aber auch außerhalb der konkreten Projektarbeit, spielt die Legitimierungsfunktion der Standards eine große Rolle. Zu Werbezwecken und für Marketingstrategien werden die Standards genauso herangezogen (Kategorie 133, 6 Nennungen), wie zum Zweck eines generellen Professionalitätsnachweises (Kategorie 132, 25 Nennungen): „Qualität der eigenen Evaluation nach außen belegen, vom Vorwurf der Willkür befreien.“ Schließlich bleibt als übergeordnetes Prinzip, die Verwendung der Standards als ethisches Schutzschild (Kategorie 135, 3 Nennungen).

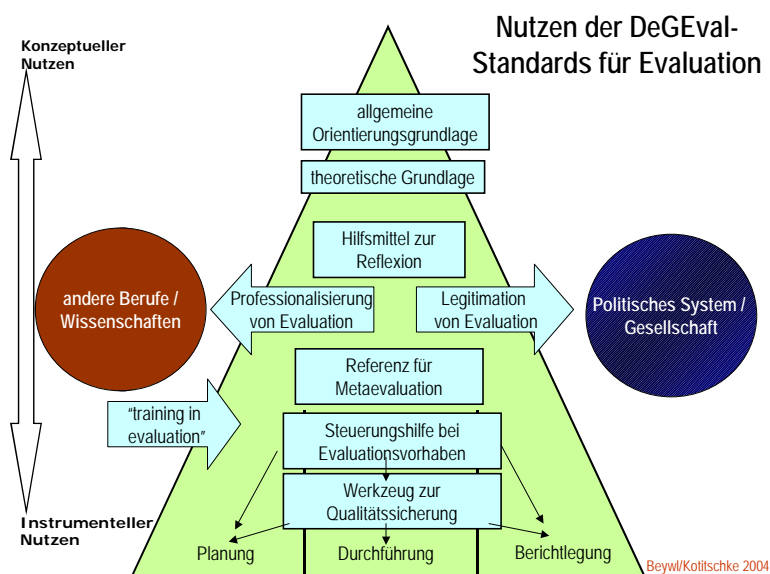


Abbildung 2: Nutzung der DeGEval-Standards für Evaluation

Ein weiteres großes Anwendungsfeld der Standards ist deren **Verwendung als Lehrgrundlage** (Metakategorie 140). Mit insgesamt 61 Nennungen in diesem Bereich umfasst diese Kategorie mit ihren 3 Untergliederungen zwar nicht so viele Antwortbeiträge wie die bereits vorgestellten Kategorien, dies ist jedoch kein Grund, die Bedeutung der Standards für Lehrzwecke gering zu schätzen. Es ist vielmehr so, dass die Nutzung der Standards zu Lehrzwecken, als noch ausbaufähige Nebenfunktion bzw. als vom direkten Verwendungszweck der Standards (Steuerungs- und Kontrollfunktion innerhalb von Evaluationsprojekten) abgehobene Funktion, Verwendung findet. Es ist anzunehmen, dass diese Nutzung der Standards zukünftig in seiner Bedeutung stetig steigen wird.

Die Funktion der Standards als Lehrgrundlage kann, je nach Differenziertheit der Antwortbeiträge, noch feiner kategorisiert werden. Neben den grundlegenden

Verwendungszwecken der Standards als Lehrgrundlage (z.B. „Einsatz in einer Lehrveranstaltung“, „kann die standards verwenden, um Evaluationsziele und -zwecke anderen zu erläutern“ oder „für die Lehre/Ausbildung in den Methoden der Evaluation“) kann eine Unterscheidung nach den Adressaten der entsprechenden Lehrbemühungen, dessen Grundlage die Standards darstellen, zu einer weitergehenden Aufklärung über die Nutzung der Standards beitragen. In diesem Sinne sammelt die Kategorie 141 die Beiträge, die explizit die Verwendung der Standards zur Einarbeitung neuer Mitarbeiter bzw. als Hilfe für Evaluationsneulinge verwenden: *Die Standards sind für meine Arbeit nützlich in der...* „Fort- und Weiterbildung von Evaluationsnovizen“. Vor allem an Universitäten kommen Sie diesbezüglich zum Einsatz:

„In meiner Vorlesung/meinem Seminar "Evaluation in der Erwachsenenbildung" weise ich die Studenten verbindlich darauf hin, dass diese Standards unumgänglich sind, falls sie eine Evaluation durchführen wollen. Die Kurzfassung aus dem Internet wird allen kopiert verteilt, auch das Buch "Joint Committee ..." in der Literaturliste mit Fettdruck hingewiesen.“

Außer für zukünftig Evaluierende, werden die Standards auch gegenüber den anderen Beteiligten innerhalb eines Evaluationsprojekts, als Lehrgrundlage verwendet. Sowohl den Auftraggebenden (Kategorie 143, 4 Nennungen), als auch den weiteren in einem Projekt Beteiligten und Betroffenen (Kategorie 142, 5 Nennungen), werden durch die Standards Grundzüge, Vorgehensweisen und Verhaltensweisen nähergebracht. Auch den Auftraggebenden sowie den Beteiligten und Betroffenen eines Evaluationsprojekts muss klar gemacht werden, welche Rolle sie im Evaluationsprozess innehaben und was sie gegebenenfalls einbringen müssen bzw. wie sie sich zu verhalten haben. Mit den Standards als Lehrgrundlage wird diesbezügliches Missverständnis vorgebeugt.

Mit den bisher dokumentierten Kategorien zur Nutzung der DeGEval-Standards sind die zahlenmäßig bedeutsamsten Verwendungszwecke vorgestellt worden. Eine Vielzahl von Antwortbeiträgen bezieht sich allerdings weniger auf einen bestimmten Zweck dem mit Hilfe der Standards dienlich sein soll, sondern mehr auf eine Nutzung der Standards in spezifischen Einsatzfeldern (Insgesamt 35 Antwortbeiträge). Diese Antworten werden in Metakategorie 150 (18 Beiträge) und seinen zwei Untergliederungen erfasst. Die beiden Untergliederungen entsprechen zwei Einsatzfeldern der Standards, die sich durch die Häufigkeit ihrer Nennung konstituieren. Zum Einen ist dies der Bildungsbereich (Kategorie 152, 11 Nennungen), und zum anderen das Einsatzgebiet der Selbstevaluation (Kategorie 151, 6 Nennungen). Andere Einsatzfelder konnten zahlenmäßig nicht hervorstechen.

Natürlich gibt es auch Antworten, die (noch) keine Nutzung der Standards benennen (Kategorie 161, 12 Nennungen), oder die Verwendung anderer Standardsets aufführen (Kategorie 160, 2 Nennungen). Die Kategorie 160 vereint Antwortbeiträge von Befragten, welche die DeGEval-Standards bisher nicht verwendet haben.

Kategorie II: Begründungen für Revisionsalternativen

KATEGORIEN	NENNUNGEN
200 – Neue Aspekte / Inhalte, die im Rahmen einer Revision der Standards eine Rolle spielen können	10
201 – Ergänzung: Fallbeispiele / Praxiserfahrung aus den verschiedenen Anwendungsbereichen der Standards	22
202 – Ergänzung: Dokumentation häufiger Fehler bei der Anwendung der Standards	6
203 – Ergänzung: Informationen und Hilfestellungen zur systematischen Anwendung der Standards (z.B. funktionales Inhaltsverzeichnis)	13
210 – Verbesserungen bestehender Aspekte / Inhalte der DeGEval Standards-Broschüre	6
211 – Optimierung: klarere Formulierung / Verständnis erleichtern (evaluators, clients, stakeholders)	8

212 – Optimierung: Erläuterungen / Definitionen an Evaluationsbedarf (in den verschiedenen Anwendungsbereichen) anpassen	9
220 – Etablierung und Professionalisierung von Evaluation und den Standards	4
221 – Begriff „Evaluation“ absichern / Professionalisierungsprozess im deutschsprachigen Raum vorantreiben	7
222 – Etablierung / Profilierung der DeGEval-Standards	6
230 – Handbuch der Deutschen Evaluationsstandards?	8
231 – umfassendes Handbuch wäre ideal	10
232 – Parallelprodukt zum Handbuch des „Joint Committee“ nicht notwendig	10
240 – Die Standards und ihre Verwendung in unterschiedlichen Anwendungsbereichen	7
241 – bessere Verbindung zwischen Theorie und Praxis	9
250 – sonstige Begründungen	21
251 – Nur eine umfassende Revision macht Sinn (keine Revision einzelner Teile der DeGEval-Broschüre)	11
252 – zunächst Anwendungserfahrungen mit den Standards sammeln	7
253 – Revision(-umfang) abhängig von stimmiger Kosten-Nutzen Relation	9
254 – Revision unter Vorbehalt – keine zwangsläufige Veränderung	8
260 – Die DeGEval-Broschüre und ihre Standards sind angemessen / bedürfen momentan keiner Revision	18
261 - keine Erfahrung mit den Standards / Unsicherheit	11

Bei der Analyse von Kategorie II („Begründungen für Revisionsalternativen“) muss die von den Befragten jeweils angegebene / begründete Revisionsalternative Berücksichtigung finden. Im Rahmen der Bearbeitung der Antwortgesamtheit mit der Text-Sortier-Technik, wurde bei den Beiträgen aus dem Textfeld von Frage 6 ein zusätzlicher Code integriert. Mit Hilfe dieser Codierung ist jederzeit ersichtlich, welche Revisionsalternative die Befragten jeweils durch Ihre Ausführungen begründen. Dieser Zusammenhang ist für eine sinnvolle Betrachtung und Interpretation der Textbeiträge essentiell.

Codes für Revisionsalternativen:

- \$1 = Revision Standardtext (25 Einzelstandards)
- \$2 = Revision Erläuterungen
- \$3 = Revision Inhalt gesamte Broschüre
- \$4 = Revision „neues Produkt“ á la Joint Committee Handbuch
- \$5 = keine Revision
- \$6 = sonstige Vorschläge für Revision
- \$7 = weiß nicht, bin unsicher

Insgesamt machen 152 Befragte (N=257) von dem freien Textfeld im Anschluss an Frage 6 Gebrauch. Die Beiträge sind in 220 Sinneinheiten aufgegliedert und im entsprechenden Kategoriensystem abgelegt worden. Im folgenden werden die häufigsten Begründungen und dazugehörige Revisionsalternativen vorgestellt.

Metakategorie 200 (10 Nennungen) unterhält mit Ihren Untergliederungen 201-203 Begründungen, die implizieren, dass sich eine Revision der DeGEval-Standards in erster Linie durch die **Aufnahme / Ergänzung neuer Aspekte / Inhalte** legitimieren würde: „Es braucht eine Erweiterung, welche die Standards und ihre theoretischen Grundlagen / die Erfahrungen, die mit ihnen gemacht würden, expliziert.“

Kategorie 201 legt dabei den Schwerpunkt auf Fallbeispiele („gute und schlechte Praxisbeispiele, kommentierte Anschauungsbeispiele geben zusätzlich Orientierungshilfen“) und die Aufnahme von Praxiserfahrungen mit den Standards („Die mittlerweile veröffentlichten und auch zum Teil im Forum diskutierten praktischen Projekte sollten in eine Revision mit eingehen.“). Die Ergänzung der bisherigen Standardpublikation (Broschüre „Standards für

Evaluation“) durch Beispiele und Anwendungserfahrungen wird in 21 Beiträgen als Grund für eine Revision angegeben.

In Kategorie 202 (6 Nennungen) steht ein anderer Aspekt im Vordergrund, der zur Begründung von Revisionsbemühungen herangezogen wird. Die Aufnahme von „häufigen Fehlern bei der Anwendung der Standards“ sollte demzufolge bei einer Revision der Standards in jedem Fall berücksichtigt werden:

„Fehler sind menschlich und können auftreten. Es ist sehr wichtig, dass Fehlerquellen im Vorfeld schon lokalisiert und behoben werden können. Aus diesem Grund ist z.B. das Themenfeld "häufige Fehler" für mich interessant.“

Schließlich wird in Kategorie 203 (13 Nennungen) die Revision damit begründet, dass Informationen und Hilfestellungen zur systematischen Anwendung der Standards, bisher in zu geringem Maße Beachtung finden. Grund genug, eine Revision in diesem Sinne auszurichten: „Die Praktikabilität der Standards sollte dokumentiert werden und künftigen EvaluatorInnen orientierende Hinweise für die eigene Arbeit geben.“

Die Begründungen der 200er Kategorien beziehen sich hauptsächlich auf die Revisionsalternativen 4 (durch 34 Sinneinheiten begründet) und 3 (durch 6 Sinneinheiten begründet). In geringer Zahl werden auch die Revisionsalternativen 1, 2 und 6 durch die Beiträge begründet.

In zwei Fällen äußern sich auch Revisionsgegner. Die Ablehnung der Revision bedeutet in diesen Fällen nicht, sich nicht doch mit deren potentiellen Inhalten auseinanderzusetzen. Man könnte von einer Ablehnung unter Vorbehalt sprechen. Ein einziges Mal in den 200er Kategorien begründet ein/e unsichere(r) Befragte(r) (Revisionsalternative 7) Ihre Zurückhaltung im Zusammenhang mit eventuellen Ergänzungen des DeGEval-Standardprodukts: „ich möchte den Aufwand für eine Revision gering halten - halte aber Anschauungsbeispiele und Richtlinien für sinnvoll - aber muss das alles in die Standards = weiss nicht, bin unsicher“

Neben den 200er Kategorien (200-203) existieren auch Begründungen, die eine Revision nicht durch die Ergänzung neuer Aspekte rechtfertigen, sondern durch die **Verbesserung / Optimierung bereits bestehender Aspekte / Inhalte der DeGEval Standards-Broschüre**. Beispielsweise könnte eine Überarbeitung der Standardtexte im Vordergrund der Revision stehen.

„Revision muss ja nicht "grundlegende Überarbeitung" bedeuten. Ich halte jedoch an einigen Stellen eine Präzisierung (sowohl der Formulierung der "Standard"-Texte als auch der Konkretheit der Erläuterungen) für sinnvoll.“

Metakategorie 210 sammelt Argumentationsansätze, die eine Revision befürworten, die sich mit einzelnen **Aspekten der aktuellen Standards** beschäftigen; sie beinhaltet zwei weitere Unterkategorien (211 und 212).

Kategorie 211 (8 Nennungen) umfasst hierbei Beiträge, die auf eine **Verständnisoptimierung der Textpassagen** abzielen: „Ein Text kann schließlich (anders als eine direkte Präsentation) nur über die verwendete Begrifflichkeit wirken; daher muss die Sprache möglichst trennscharf und un-missverständlich sein.“

Bestehende Formulierungen sollen dahingehend optimiert werden, dass sie für Evaluierende (ob erfahren oder unerfahren), als auch für Auftraggebende und Beteiligte / Betroffene leichter verständlich sind:

„Die Wichtigkeit und Relevanz und damit auch Anwendung einzelner Standards oder Richtlinien dürfte abhängig davon sein, ob man es mit Evaluatoren, Auftraggebern, Betroffenen etc. zu tun hat. Das bedeutet nicht, dass man für alle Gruppen unterschiedliche Standards braucht; das

bedeutet aber sehr wohl, dass die unterschiedlichen Ansichten und Perspektiven zu berücksichtigen sind“

Kategorie 212 (9 Nennungen) bezieht sich konkret auf die Optimierung der Erläuterungen / Definitionen. Hierbei spielt vor allem die Übertragbarkeit auf verschiedene Anwendungsbereiche eine Rolle:

„Es fehlen noch Kriterien für bestimmte Anwendungsbereiche bzw. es sind je nach Anwendungsbereich Kriterien zu relativieren. Daher muss sich die Revision sowohl auf die Einzelstandards als auch auf die Erläuterungen dazu beziehen.“

Die Begründungen, die sich in den 210er Kategorien (210-212) wiederfinden (insgesamt 23 Nennungen), werden erwartungsgemäß zur Legitimation einer Revision nach dem Vorbild von Alternative 3 (durch 9 Sinneinheiten begründet) herangezogen. 10 weitere Begründungsbeiträge beziehen sich auf die Alternativen 1, 2 und 4.

Auch ein/e Revisionsgegner/in äußert sich im Rahmen der Thematik von Kategorie 210. Wie bereits oben erläutert handelt es sich auch hier um eine Äußerung unter Vorbehalt („Falls es aus bestimmten Gruppen/Praxisbereichen Hinweise gibt, dass einzelne Aspekte hinzugefügt oder überarbeitet werden müssten, sollte das selbstverständlich geschehen.“).

Mit Metakategorie 220 und ihren 2 Untergliederungen (221 und 222) existiert ein weiteres Sammelbecken für mögliche Begründungen. In diesem Fall werden Beiträge aufgeführt, die mit einer Revision die **Etablierung und Professionalisierung von Evaluation und den Standards** in Verbindung bringen. Eine Revision könnte hierzu neue Impulse setzen und Bedingungen schaffen, die eine interdisziplinäre Verbreitung und Weiterentwicklung von Evaluation und den Standards unterstützen. Die Autoren der Beiträge in dieser Kategorie sehen konkreten Handlungsbedarf, um diesen Prozess voranzutreiben: „Die Verbreitung der Standards - insbesondere hinsichtlich der praktischen Nutzung bei Auftraggebern und Auftragnehmern - ist nach meiner Einschätzung alles andere als ein Selbstläufer.“

In Kategorie 221 werden die Beiträge gesammelt, die Evaluation explizit als Begriff und Methode absichern wollen, und hierin den Grund für eine Revision sehen:

„...den Begriff Evaluation bzw. die Qualität durchgeführter Evaluationen erheblich anheben bzw. "absichern" und dadurch sowohl auf Seiten der EvaluatorInnen als auch auf Seiten der Auftraggeber, aber auch der Beteiligten, zu mehr "Sicherheit", "Klarheit" und auch "Vertrauen" führen.“

Evaluation als Begriff und Methode abzusichern und im deutschsprachigen Raum zu etablieren gelingt nur im Einklang mit überzeugenden Standards. Diese Erkenntnis schlägt sich in den Beiträgen in Kategorie 222 (Etablierung / Profilierung der DeGEval-Standards) nieder: „Ein Ausdifferenzierung der Standards für Deutschland halte ich für sehr hilfreich, um den Standards in Deutschland mehr Gewicht zu verleihen.“. Die Etablierung und Profilierung der DeGEval-Standards wird dort konkret als Begründung für anstehende Revisionsbemühungen herangezogen. Die DeGEval-Standards sollen ein eigenständiges Profil erhalten, angepasst an Bedarfe und Bedingungen im deutschsprachigen Raum. In diesem Zusammenhang wird oft eine Abgrenzung zu den Standards des Joint Committee⁴ thematisiert („die Standards sollten gegenüber den

⁴ Das Joint Committee on Standards for Educational Evaluation ist eine Non-Profit-Organisation zur Entwicklung und Überprüfung von Standards für Evaluation in Bildung und Erziehung. 1975 durch die American Educational Research Association, die American Psychological Association und dem National Council on Measurement in Education ins Leben gerufen, wurde ihre Mitgliedschaft früh durch andere Organisationen ausgeweitet, um weitestgehend allgemein geltende Standards zu generieren.

Standards des Joint Committees ein eigenständiges Profil erhalten. dies könnte gelingen durch eine zusätzliche Version“). Das „Handbuch der Evaluationsstandards“⁵ wird innerhalb dieser Kategorie häufig als Maßstab oder Vorlage angesehen. Ein vergleichbares Produkt für den deutschsprachigen Raum würde die Etablierung und Professionalisierung von Evaluation und den DeGEval-Standards nach Einschätzung einiger Befragungsteilnehmenden immens vorantreiben:

„Ein dem JC-Handbuch vergleichbares von der DeGEval herausgegebene Produkt für den deutschsprachigen Raum würde m.E. für Deutschland über die jetzigen Standards hinausgehende Maßstäbe setzen (insbesondere im Hinblick auf Umsetzungen und Interpretation)“

Es ist deswegen auch nicht verwunderlich, dass Begründungen der Kategorie 220 (220-222) mit großer Mehrheit bei einer Entscheidung für Revisionsalternative 4 Verwendung finden. 13 der 17 Beiträge begründen die Wahl von Alternative 4. Je ein Mal wird Alternative 2 und Alternative 3 begründet.

Auch in dieser Kategorie finden sich zwei Befragungsteilnehmende, die eine Revision ablehnen. In einem Fall kann auch hier von einer Ablehnung unter Vorbehalt gesprochen werden. Der Vorbehalt bezieht sich auf die nach der Aussage des Autors fehlenden Ressourcen, um eine umfassende Revision (Alternative 4) in absehbarer Zeit leisten zu können:

„Generell halte ich eine Buchfassung im Stile des Joint Committee für sinnvoll, allerdings sehe ich derzeit, d.h. in Anbetracht der nächsten 1-2 Jahre, nicht die Möglichkeit, unter den Mitgliedern der DeGEval die Ressourcen zu mobilisieren, ein solches Kompendium zu erstellen“.

Im anderen Fall macht die Entscheidung für Revisionsalternative 5 angesichts des Antwortbeitrags nur wenig Sinn („die Standards sollten noch weiterläufiger etabliert werden“). Der Befragungsteilnehmende fordert eine „weitläufige Etablierung“ ohne eine Revision anzustreben.

Das bereits genannte „**Handbuch der Evaluationsstandards**“ spielt bei den Überlegungen der Befragungsteilnehmenden bezüglich einer Revision der DeGEval-Standards eine große Rolle. Die Frage, ob auch die DeGEval ein derartiges Produkt entwickeln sollte, wird vielfach kommentiert und unterschiedlich ausgelegt. Die große Anzahl an Textbeiträgen zu diesem Thema legt nahe sie explizit zu kategorisieren. Metakategorie 230 (230-232) umfasst somit hauptsächlich Textbeiträge, die sich ausdrücklich für (231) oder gegen (232) eine Revision nach Alternative 4 aussprechen. Oberkategorie 230 enthält Äußerungen, die in Ihrer Zustimmung / Ablehnung Interpretationsspielraum offen lassen bzw. Vorbehalte ausdrücken:

„Wenn überhaupt eine Revision vorgenommen wird (was ich persönlich nicht für absolut dringend halte), dann sollte sie umfassend sein und das Werk des Joint Committee ist eben doch hilfreicher als eine Kurzfassung.“

„eine nachdrückliche Förderung systematischer und vor allem auch methodisch zumindest halbwegs korrekter Evaluation durch Nicht-Spezialisten in Organisationen würde zwangsläufig auf die Option (4) hinauslaufen.“

„Ich persönlich könnte mir gut eine Revision unter (4) vorstellen, weil ich die Standards in diesem Sinne als "Handbuch" verwende. Ich komme aber auch gut mit der Buchfassung des "Joint Committee" klar.“

⁵ Joint Committee on Standards for Educational Evaluation: Handbuch der Evaluationsstandards, hg. von James R. Sanders. 2. durchgesehene Aufl., Opladen 2000.

Deutlicher sind die Beiträge aus Kategorie 231 (11 Nennungen). Hier wird die Entscheidung für Revisionsalternative 4 begründet und somit ein Handbuch der deutschen Evaluationsstandards als ideales Ziel einer Revision propagiert:

„Halte ich für die Optimallösung (falls leistbar). Bei der Vielschichtigkeit von Evaluierungen wäre dies auch und gerade für Auftraggeber eine sinnvolle Arbeitshilfe.“

„Gerade weil es wenig ausgewiesene Experten mit Erfahrung und entsprechendem theoretischen Wissen gibt, braucht es ausführliche, gut gemachte Publikationen.“

„Auch wenn es vermutlich die Arbeitskapazitäten der Beteiligten sprengen mag, so wäre doch eine Fassung a la "Joint Committee" mit Beispielen aus verschiedenen Politikfeldern hilfreich, da man im Rahmen der Übersetzung der Standards, die notwendigerweise auf einem hohem Abstraktionsniveau formuliert wurden, auf die Anwendungsprobleme vor Ort immer auf Dilemmata stößt und die Vermittlung zwischen den verschiedenen Ansprüchen, die durch die Standards markiert werden, das eigentliche Problem der Praktiker darstellen, die dieses Buch ja schließlich benutzen sollen.“

Es wird deutlich, dass Alternative 4 neben der in Kategorie 220 aufgeführten Funktion der Etablierung und Professionalisierung der Standards, aus vielerlei weiteren Gründen für die Befragungsteilnehmenden attraktiv erscheint. Dabei wird jedoch nicht verkannt, dass die Realisierung eines solchen Vorhabens mit immensem Aufwand verbunden ist.

Bei den Textbeiträgen von Kategorie 232 haben sich erwartungsgemäß keine Personen geäußert, die sich für Alternative 4 entschieden haben.

Das ein Parallelprodukt zum Handbuch des Joint Committee momentan nicht notwendig ist, wird von 10 Personen als Begründung für Ihre Alternativenauswahl herangezogen. Dabei lehnen 5 Personen eine Revision ab (Alternative 5). 5 weitere Personen begründen damit die Wahl für Revisionsalternative 2 (1 mal), 3 (3 mal) oder 6 (1 mal).

„derzeitige Stärke der DeGEval Broschüre ist seine Kürze und Prägnanz. Eine Ausweitung, wie unter Punkt 4 angesprochen, erscheint unnützlich, weil es das ja schon gibt.“

„Die bestehende Version scheint mir praxisbezogen und daher ihrem Zweck entsprechend. Für detailliertere Ausführungen stehen die Standards des Joint Committee zur Verfügung. Warum sollte die DeGEval hier ein "Konkurrenzwerk" schaffen?“

„ich finde die Standards im großen und ganzen gut, sie können als allgemeine Leitlinien gelten (als "Rahmen") - natürlich sind sie recht allgemein und abstrakt formuliert, eine Anpassung an den jeweiligen Kontext muss selbst vorgenommen werden - das finde ich aber gut“

Die Anpassung der Standards an unterschiedliche Kontexte bzw. die **Verwendung der Standards in unterschiedlichen Anwendungsbereichen** stellt einen weiteren Aspekt dar, der im Rahmen der Begründung für eine bestimmte Revisionsalternative Verwendung findet.

Wie das letztgenannte Zitat eines Teilnehmenden verdeutlicht, werden die Standards einerseits als allgemeine Leitlinie angesehen, um - ohne konkrete Handlungsanweisungen aufzuführen - eine Struktur anzubieten, die in Evaluationen Orientierungspunkte setzt. Andererseits wird von den Standards eine spezielle Ausrichtung und Anpassung für verschiedene Einsatzfelder gefordert, die Besonderheiten in den jeweiligen Feldern berücksichtigt.⁶

Metakategorie 240 (7 Nennungen) sammelt Beiträge, die im Rahmen der Revision hinsichtlich der praktischen Verwendbarkeit der Standards in unterschiedlichen Bereichen, eine Auseinandersetzung anstreben: „an den spezifischen Anforderungen

⁶ Die dargestellte Ambivalenz wird im Kapitel „Issues“ weitergehend beleuchtet werden.

einzelner Fachbereiche im Bezug auf die Evaluationsstandards sollte weiter gearbeitet werden, um daraus ggf. die Bedarfe einer Revision abzuleiten“. **Gegenstand einer Revision wäre demzufolge die Entwicklung und Analyse möglicher Differenzierungen, welche die Standards zur Verwendung in verschiedenen Bereichen optimiert. Gleichzeitig könnte die Revision klären, inwiefern die Standards bereits das Potential aufweisen, in verschiedenen Evaluationsfeldern spezifisch wirksam zu sein. Gegebenenfalls sollten die Standards stärker dahingehend ausgerichtet werden: „Es wäre m.E. zu klären inwiefern die Standards auf unterschiedliche Felder der Evaluationsforschung anwendbar sind. Möglicherweise ließen sich hier Teilmengen bilden.“**

Kategorie 241 (9 Nennungen) geht ganz konkret auf die tatsächliche Übertragung der „theoretischen Standards“ in die praktische Realität ein. Einige Befragungsteilnehmende sehen im Theorie-Praxis-Problem den entscheidenden Grund für eine Revision:

„Mir erscheint die Implementierung in die Praxis derzeit wichtiger als eine detailreichere Differenzierung aufgrund theoretischer Diskussionen“

„Art und Umfang muss sich im Rahmen der Arbeit verwirklichen lassen auf der Grundlage, dass es sich um ein Praxisfeld handelt und nicht durch unüberprüfbare Konzepte ersetzt wird.“

„Eine Revision sollte sich stärker an einer praxisnahen Überarbeitung orientieren, d. h. Standards als Diskussionsgrundlage sind ja gut und schön, doch ich würde eine stärkere praxisnahe Orientierung bevorzugen.“

Metakategorie 240 vereint Begründungen für alle im Online-Fragebogen vorgeschlagenen Revisionsalternativen. Keine Revisionsalternative wird in besonders häufigem Maße auf diese Weise begründet.

Ein Beitrag dieser Kategorie stammt von einer Person, die eine Revision momentan ablehnt. Die Person begründet Ihre Ablehnung damit, dass momentan in erster Linie die Übertragung und Etablierung der Standards in die Praxis vorangetrieben werden sollte.

Die Aufforderung, die Wahl für eine bestimmte Revisionsalternative zu begründen, erbrachte eine Vielzahl unterschiedlichster Argumente. In den bisher aufgeführten Kategorien wurden die Begründungen zusammengefasst, welche besonders häufig und eindringlich dargestellt wurden. Neben diesen Hauptargumentationslinien wird eine zukünftige Revision in Abhängigkeit zu weiteren Aspekten befürwortet oder abgelehnt. Metakategorie 250 sammelt all die Begründungen, welche thematisch nicht in die bereits bestehenden Kategorien integriert werden konnten (Metakategorie 250: „**sonstige Begründungen**“).

Kategorie 250 beinhaltet 4 Untergliederungen. Diese konstituieren sich durch besonders häufig genannte Aspekte zur Begründung (oder Ablehnung) einer Revision.

Kategorie 251 (11 Nennungen) enthält Beiträge, die eine Revision nur dann befürworten, wenn Sie umfassend ist und alle Bereiche und Aspekte der Standards und der Standardsbroschüre thematisiert. Eine Revision einzelner Teile, wie beispielsweise die Revision der Erläuterungen in der Standardsbroschüre (Revisionsalternative 2 des Online-Fragebogens) wird abgelehnt: „Standards mit ihren Erläuterungen sowie Einführung einschl. Definitionen u.a. sind im Zusammenhang zu sehen und deshalb bei einer Revision insgesamt zu überprüfen und ggf. zu überarbeiten.“

Es ist nicht verwunderlich das die Autoren der Beiträge in Kategorie 251 ausschließlich Revisionsalternative 3 und 4 begründen. Diese Alternativen entsprechen den Vorgaben einer umfassenden Revision.

Die Textbeiträge in Kategorie 252 (7 Nennungen) drücken einen weiteren Vorbehalt aus, welcher mit der Durchführung einer Revision verbunden ist. Eine Revision soll demzufolge erst dann verwirklicht werden, wenn weitere Anwendungserfahrungen mit den Standards als Datenbasis für eine Revision vorliegen: „Ich finde, man sollte Revisionen nur in größeren Zeitabständen vornehmen und erst mal noch eine Weile Reaktionen und Anwendungserfahrungen sammeln. Man rennt sonst nur einer immer vorläufigen Aktualität hinterher.“ Dass zunächst ein empirisches Fundament geschaffen werden sollte, auf dem eine Revision aufbauen könnte, drückt auch folgendes Zitat eines Teilnehmenden aus.

„Die Revision der Standards sollte gründlich sein, um sicherzustellen, dass sie eine möglichst hohe Akzeptanz erhalten - und nicht an dieser Stelle Angriffsfläche bietet. Selbst als jemand, der sehr viel mit den Standards arbeitet - oder gerade deshalb - sehe ich dringenden Revisionsbedarf in den Standards und in den Erläuterungen. Der Revisionsprozess soll aber nicht übereilt werden, da m.E. Basis für den Prozess eine breite Anwendung und praktische Erfahrung mit den Standards sein sollte. Der Revisionsprozess sollte möglichst auch empirisch fundiert sein. Ich plädiere für eine Verlängerung der Frist für den Revisionsprozess.“

Teilnehmende der Befragung begründeten auf diese Weise in Erster Linie ihre Ablehnung gegenüber einer Revision (Revisionsalternative 5). Auch Revisionsalternative 6 und 7 wird mit diesem Vorbehalt begründet. Es ist demnach nicht der Fall, dass ausschließlich Revisionsablehner ihre Wahl derart begründen. Auch unsichere Personen bzw. Personen mit alternativen Vorstellungen einer Revision (im Vgl. zu den vorgeschlagenen Alternativen des Online-Fragebogens) begründen ihre Entscheidung auf diese Weise. Bei einer Person konnte nicht hergeleitet werden, welche Revisionsalternative gewählt wurde. Der Antwortbeitrag wurde aus einer anderen Kategorie importiert, die nicht mit der Codierung für eine Revisionsalternative versehen ist.

In Unterkategorie 253 (9 Nennungen) sind Beiträge enthalten, welche die Zustimmung für eine Revision von einer stimmigen Kosten-Nutzen Relation abhängig machen: „Besteht eine sinnvolle Kosten-Nutzen-Relation?“. Dieser Vorbehalt wurde bereits in den Beiträgen von Kategorie 230-231 deutlich. Dort wird eine Revision thematisiert, die umfassend ist und ein Produkt in Form eines Handbuchs hervorbringt. Falls die Ressourcen für ein derartiges Vorhaben vorhanden sind wäre auch der Nutzen entsprechend hoch und die Revision zu befürworten:

„Es braucht eine Erweiterung, welche die Standards und ihre theoretischen Grundlagen / die Erfahrungen, die mit ihnen gemacht wurden, expliziert; das verlangt eine grosse Anstrengung, Mittel und Personeneinsatz, würde sich dann allerdings lohnen.“

Die Begründungen dieser Kategorie vereinen Befragte mit gegensätzlichen Entscheidungen hinsichtlich der Wahl für eine Revisionsalternative. Sowohl Revisionsablehner, als auch die Verfechter einer umfassenden Revision mit neuem Produkt beziehen sich auf diesen Zusammenhang, um ihr Votum zu begründen.

Schließlich werden in Kategorie 254 (8 Nennungen) alle übrigen Begründungen aufgelistet, bei welcher eine Revision nicht zwangsläufig mit weitreichenden Veränderungen einhergehen muss, d.h. nur unter Vorbehalt stattfindet: „Aufgabe der Revision sollte es sein, die bislang gesammelten Erfahrungen mit den Standards auszuwerten und zu prüfen, ob dies zur Weiterentwicklung der Standards Anlass gibt“

Beiträge dieser Kategorie weisen keine besondere Struktur bezüglich einer bestimmten Revisionsalternative auf, die durch sie begründet wird. Sowohl Revisionsablehner (2 mal), als auch Befürworter einer Revision der gesamten Standards-Broschüre (3 mal) tauchen in dieser Kategorie auf. Zudem 2 Personen, die im Rahmen einer Revision ausschließlich den Standardtext berücksichtigen würden, sowie eine Person die ein neues Produkt als Ergebnis der Revision

bevorzugen würde. Diese Zusammensetzung ist zwangsläufig auf die thematische Offenheit der Kategorie zurückzuführen.

Dahingegen enthält die letzte Metakategorie hauptsächlich Beiträge zur Begründung für die Wahl von Revisionsalternative 5 und 7. Metakategorie 260 (18 Nennungen) enthält Textbeiträge, welche die DeGEval-Broschüre und ihre Standards für angemessen halten und deswegen **momentan keine Revision propagieren**.

„Ich finde die Broschüre gelungen strukturiert und die Sprache verständlich.“

„Broschüre ist knapp und präzise, wenn auch nicht allzu detailliert. Aber dies ist kein wesentlicher Mangel, denn sie soll ja Leitlinien aufzeigen und keine Evaluationskonzepte für einzelne Maßnahmen oder Projekte.“

„Die derzeitigen DeGEval-Standards genügen meinen eigenen Qualitätsansprüchen und Arbeitserfordernissen. Ich sehe im Moment auch keine wesentliche Weiterentwicklung auf internationaler Ebene, bspw. bei European Commission: Evaluating EU Activities (Nov. 2003).“

Unterkategorie 261 (11 Nennungen) gliedert die Ablehnung gegenüber einer Revision nach einem anderen Schwerpunkt. In diesem Fall konstituiert sie sich durch mangelnde Erfahrung / Kenntnis mit den Standards: „Da ich die Standards i.w. nur "theoretisch" kenne (also nicht aus Anwendung und Umsetzung), kann ich nicht unbedingt beurteilen, wo eine Revision notwendig oder sinnvoll sein könnte“.

Kategorie III: Kommentare / Anmerkungen zu den Standards und dem Revisionsprozess

KATEGORIE	NENNUNGEN
300 – formale Verbesserungsvorschläge	8
301 – Handbuchpublikation der DeGEval	9
302 – Gliederung der Standards	5
303 – verschiedene Publikationsformen der Standards	10
310 – inhaltliche Verbesserungs-, und Erweiterungsvorschläge	37
311 – Fallbeispiele (aus verschiedenen Anwendungsbereichen)	26
312 – Hinweise und Hilfestellungen zur Anwendung der Standards	20
313 – Dokumentation von häufig gemachten Fehlern bei der Anwendung der Standards; Problemdarstellung	13
314 – Aufnahme von Lösungsansätzen bei Problemen bezügl. der Anwendung der Standards	6
315 – Erläuterungen zur Anwendung der Standards in verschiedenen Kontexten	18
316 – Erweiterung zur Anwendung der Standards im Rahmen bestimmter Evaluationsformen (Fremd- und Selbstevaluation, formativ / summativ, Metaevaluation)	15
317 – Erweiterung zur Interaktion zwischen Evaluierenden, Evaluierten und Auftraggebenden	9
318 – Darstellung von Zusammenhängen zum Verständnis einer Evaluation und deren Wirkungsweise	6
320 – weniger wissenschaftliche / theorielastige, stattdessen praxisnahe Präsentation der Standards	12
330 – Kommentare / Anmerkungen zu den Standards und / oder den Begleittexten, Erläuterungen	14
331 – positive Kommentare / Anmerkungen	13
332 – negative Kommentare / Anmerkungen	18
333 – neuer Standard	2
334 – Kommentare / Anmerkungen zur Einleitung	2
335 – Kommentare / Anmerkungen zu den Erläuterungen	6
336 – Kommentare / Anmerkungen zu den Standardgruppen	5
337 – Kommentare / Anmerkungen zu einzelnen Standards	7
337A – Standardname	2
337B – Standardtext	7
337C – Standarderläuterungen	1

338 – Kommentare / Anmerkungen zum funktionalen Inhaltsverzeichnis	2
339 – Kommentare / Anmerkungen zu sonstigen Einzelbestandteilen	1
340 – Kommentare / Anmerkungen zum Revisionsprozess	15
341 – konkrete Vorschläge zum Ablauf einer Revision	14
342 – skeptische / ablehnende Haltung gegenüber einer Revision	6
350 – sonstige Bemerkungen	16

Wie bereits erwähnt, handelt es sich bei Kategorie III um die Antwortgesamtheit zwei offener Fragen mit ähnlichem Antwortpotenzial. Zum einen Frage 8 des Online-Fragebogens („Was könnte Ihrer Auffassung nach verbessert werden? Welche Veränderung / Erweiterung / Verbesserung schlagen Sie vor?“), zum anderen Frage 15 („Evtl. Anmerkungen/Kommentare zum Thema Standards/Revisionsprozess“). Zusammen ergeben sich 139 Textbeiträge, die in 331 Sinneinheiten aufgegliedert worden sind. Das Kategoriensystem, in welchem die Antworten abgelegt worden sind, enthält sechs Metakategorien, die wiederum in insgesamt 26 Untergliederungen aufgeteilt sind. Die feine Ausdifferenzierung dieses Kategoriensystems macht es möglich, die Antworten zweier unterschiedlicher Fragen vereinigend aufzunehmen, ohne das dabei die Aussagekraft hinsichtlich des intendierten Zwecks der beiden Fragen verloren geht. Die Zusammenführung war dennoch nur möglich, da das Profil der Antwortgesamtheit beider Fragen in sehr vielen Punkten homogen ist, d.h. ähnliche Zusammenhänge / Argumentationsstränge, Analysen und Vorschläge einschließt.

Bei der Aufteilung der Verbesserungsvorschläge wurde, wie dies auch von Teilnehmenden vorgeschlagen worden ist, zunächst in formale und inhaltliche Vorschläge unterschieden:

„Verbesserungen der Standards würde ich in zwei Klassen einteilen. Formale Veränderungen / Überarbeitungen beziehen sich eher auf den Ausdruck, den Stil, die Verwendung von Sprache. So sollte das Redaktionsteam m. E. beispielsweise einige überlange Sätze in kürzere Sätze teilen, Substantivierungen in flüssigere Verbalformen umwandeln oder Passivkonstruktionen, so weit es geht, aus dem Text herausnehmen. Inhaltliche Änderungen beziehen sich eher auf den Gehalt, auf die Sache.“

Zunächst werden in Metakategorie 300 und seinen drei Untergliederungen **formale Verbesserungsvorschläge** zur Optimierung der Standards gesammelt.

Metakategorie 310 dokumentiert anschließend die inhaltlichen Verbesserungsvorschläge der Befragungsteilnehmenden. Dabei finden acht Unterkategorien Verwendung. Diese feine Untergliederung ermöglicht eine differenzierte Betrachtung der inhaltlichen Vorschläge.

Die drei Untergliederungen von Kategorie 300 stellen die am häufigsten genannten formalen Aspekte dar, die im Zusammenhang mit der Verbesserung und Erweiterung der Standards eine Rolle spielen. Kategorie 301 (neun Nennungen) sammelt hierbei sämtliche Vorschläge, die auf die Entwicklung einer mit dem Standards-Handbuch des Joint Committee vergleichbaren Publikation der DeGEval abzielen, und darin die entscheidende Erweiterung der DeGEval-Standards sehen:

„Ich würde die Arbeit auf die Erweiterung im Sinne einer Buchfassung a la Joint Committee konzentrieren.“

„Mittel- bzw. langfristig halte ich die Entwicklung eines Handbuchs (analog Joint Committee) für hilfreich“

In diesem Zusammenhang werden von einigen Teilnehmenden bereits Wege aufgezeigt, auf welche Weise eine derartige Publikationsform realisiert werden sollte:

„Entwicklung eines Buches vergleichbar dem Handbuch der Evaluationsstandards; Einbezug einer großen Zahl - z.B. von ca. 100 Fachpersonen - auch über die unmittelbare DeGEval-Mitgliedschaft hinaus.“

Neben einer Publikationsform nach Vorbild des Joint Committee Handbuch werden in Kategorie 303 (zehn Nennungen) weitere Publikationsformen genannt, die von den Teilnehmenden als Bereicherung / Erweiterung hinsichtlich einer gezielten und vermehrten Verwendung der Standards empfunden würden. In dieser Hinsicht wird beispielsweise eine Publikationsform der Standards gefordert, die sich am Ablauf einer Evaluation orientiert:

„eine zusätzliche Version (der Standards), die sich am Evaluierungsablauf orientiert, indem Fragen der institutionellen Verankerung einbezogen werden sowie die Nutzung von und das Lernen aus Evaluation ausdifferenziert wird.“

Weiterhin werden Kurzfassungen (u.a. für Auftraggebende), Handreichungen und Hilfsmittel (z.B. interaktive Version) sowie Publikationsvarianten der Standards für bestimmte Anwendungsbereiche angedacht.

„Vielleicht sind 2 Versionen der Kurzfassung sinnvoll. Einmal alle 25 Standards für die Selbst- und Fremdkontrolle und eine Minimal-Version für die Hand der (eiligen) Auftraggeber.“

„Handreichungen bzw. Hilfsmittel in Form von Arbeitsmaterialien evtl. in Softwareformat (z.B. pdf-Datei) könnten zu den einzelnen Standards entwickelt werden.“

„interaktive Form, die Nachschlagen ermöglicht“

Ähnliche Vorschläge finden sich in Kategorie 302 (5 Nennungen). Dort werden jedoch nicht weitere Publikationsformen angestrebt, sondern die bisherige Gliederung der Standards in den Fokus möglicher Verbesserung gerückt. Auch in diesem Zusammenhang wird genannt, dass der Ablauf einer Evaluation als strukturelle Vorgabe für eine benutzerfreundlichere Gliederung Berücksichtigung finden könnte: „Organisation der Standards nach dem Ablauf einer Evaluation gliedern, dann sind sie besser einsetzbar.“

Insgesamt wird es als hilfreich angesehen, die Standards auch nach Anwendungsaspekten zu gliedern:

„es wäre hilfreich, die Standards auch nach eher praktischen Gesichtspunkten gruppiert zu haben, auch die von König vorgeschlagenen Kriterien verdienen weiterer Überlegung (Zeitschrift für Evaluation, 1/2003, S. 87)“

Metakategorie 300 (8 Nennungen) umfasst all die formalen Veränderungsvorschläge, die aufgrund ihrer allgemeinen Ausrichtung oder seltenen Nennung in keiner der themenspezifische Kategorien 301-303 abgelegt sind. Die Vorschläge / Anmerkungen beziehen sich auf die Formulierung der Standards (z.B. „zu lange und manchmal umständliche Formulierung der Standards. Die könnte man abspecken. In der Hälfte des Textes könnte man das Gleiche sagen.“), eine mittelfristig anzustrebende Zertifizierung über den Umgang mit den Standards, ein anzufügendes Glossar und die Anmerkung, dass die Standards auch auf Französisch angeboten werden sollten.

In der anschließenden Metakategorie 310 werden **inhaltliche Verbesserungsvorschläge** aufgeführt. Kategorie 310 vereint mit seinen neun Untergliederungen (311-318) 157 Textbeiträge bzw. Sinneinheiten. Damit ist sie mit Abstand die größte Kategorie innerhalb des Kategoriensystems III. Kategorie 310 (37 Nennungen) enthält Beiträge, die vereinzelt genannte Erweiterungsaspekte darstellen oder zu allgemein formuliert sind, um in eine der thematischen Untergliederungen (311-318) aufgenommen zu werden. Exemplarisch werden im Folgenden einige dieser Aspekte aufgeführt:

Gender Mainstreaming berücksichtigen:

„Eine Ergänzung der Standards im Sinne des Gender Mainstreaming ist ebenfalls unabdingbar.“

„Clearingstelle“ für Streitfragen einrichten:

„Ich halte vor allem die Einrichtung einer "Clearingstelle" zur Klärung von Streitfragen bezüglich der Auslegung und Einhaltung von Standards für dringend erforderlich. Ohne eine solche Stelle werden sich die Standards auf Dauer nicht als solche durchsetzen. Dies gilt verstärkt, wenn nun auch noch Standards zur Selbstevaluation verabschiedet werden, auf die sich Organisationen eventuell berufen, ohne sie jedoch zu beachten!“

deutlicherer Einbezug verschiedener Perspektiven:

„Einbezug der unterschiedlichen Perspektiven von verschiedenen, an einem Evaluationsprozess beteiligten Personengruppen“

Kriterien zur Beurteilung von Evaluationen aufnehmen:

„Die Standards sollten zumindest auch Standardkriterien der Beurteilung (vgl. DAC) beinhalten, einen Minimalkatalog und eine erweiterte Liste.“

**vergleichende Analyse
(Joint Committee Standards, SEVAL Standards, ...):**

„vergleichende Analyse“, „Herausfiltern der (feinen) Unterschiede zu den SEVAL-Standards“

Zusammenrücken: Evaluation / Standards in Europa:

„Würde mir ein Eingehen auf andere Standards wünschen, damit Europa hier mehr zusammenfinden kann und nicht nur jedes westeuropäische Land und jede Organisation mit eigenen Standards auftritt. Also Einbindung in eine europäische Diskussion mit dem Ziel, eine Annäherung zu erreichen, um wenigstens in Richtung Osten die Vielfalt der Standards etwas zu vereinfachen.“

„Sich mit Österreich und Schweiz auf einen Standard einigen. Wenn die Standards wirklich gut sind und die wesentlichen Elemente einer Evaluation enthalten, dann sollte es keinen Unterschied geben.“

Im Gegensatz dazu entsprechen die Beiträge in den Unterkategorien 311-318 besonders häufig genannten Aspekten zur inhaltlichen Erweiterung der Standards. Mit 26 Nennungen stellt Kategorie 311 die umfangreichste Untergliederung dar. In ihr werden die Beiträge aufgelistet, die eine Erweiterung der Standards um Fallbeispiele vorschlagen:

Erweiterung um Fallbeispiele
„Zusätzlich fände ich pro Hauptstandard 3-4 Praxisbeispiele gut - hier sollte nur ein Verweis auf eine entsprechende URL oder Publikation gemacht werden, um weiter forschen zu können - vielleicht könnten interessierte Praxisprojekte jeweils 2-3 Seiten zu ihren Erfahrungen, Chancen und auch Problemen in der Praxis zu dem entsprechenden Standard anbieten.“
„Vor allem auf Probleme, die die Praxis mit sich bringt, anhand von Beispielen, hinweisen. Beispiele erhöhen das Verständnis des einzelnen Standards, wie der Verwendung des gesamten Standards-Sets.“
„Erweiterung um Fallbeispiele, in denen konkrete Entscheidungssituationen aus der Evaluationspraxis anhand der Standards diskutiert werden.“
„Noch sehr "theorielastig" - in vielen Bereichen fehlen einfache und vor allem "unakademische" Beispiele.“

Kategorie 312, mit 20 Nennungen zweitgrößte Untergliederung der Metakategorie „inhaltliche Verbesserungsvorschläge“, listet Beiträge auf, die Hinweise und Hilfestellungen zur Anwendung der Standards geben. Die Beiträge variieren in der Art, wie konkret die Vorschläge formuliert sind. Einerseits wird ganz generell gefordert, dass Hinweise und Hilfestellungen zur Anwendung der Standards aufgenommen werden sollen:

Hinweise und Hilfestellungen zur Anwendung der Standards
„eine Mischung aus konkreten Hinweisen und Hilfestellungen zur Anwendung der Standards“
„Richtlinien für die Anwendung eines Standards“
„Die Standards sollten weiter erläutert und mit einfachen Instrumenten zur Vorgehensweise untermauert werden.“

Andererseits wird auch konkret formuliert, in welchen Zusammenhängen Hinweise und Hilfestellungen besonders notwendig sind:

Hinweise und Hilfestellungen II
„Der Umgang mit sich gegenseitig ausschließenden Standards sollte vertieft werden. Ein beispielhaftes Verfahren bzw. Bausteine könnten hilfreich sein“
„wobei die Grenzen der einzelnen Standards allerdings auch aufgeführt werden sollten, d.h. bis zu welchem Punkt in der konkreten Umsetzung der Standards ist Spielraum möglich, ab welchem Punkt ist eine "saubere" Evaluation dagegen nicht mehr gegeben.“

In Kategorie 313 werden Verbesserungsvorschläge gesammelt, die nahe legen, häufig gemachte Fehler bei der Anwendung der Standards zu dokumentieren. Eine Ergänzung der Standards um eine ausführliche Problemdarstellung im Umgang mit den Standards steht bei den 13 Beiträgen dieser Kategorie im Mittelpunkt:

Erweiterung um häufig gemachte Fehler
„Beispiele für häufige Fehler“
„auf häufige Fehler/Missverständnisse hinweisen“
„Verweis auf mögliche Fehler“
„verstärkt auf die verbreitetsten Fehler und deren Auswirkungen sowie auf gelungene Beispiele - evtl. mit Querverweis zu Homepages etc. - eingehen.“

Mit sechs Nennungen wird auch der Vorschlag unterstützt zusätzlich zur Problemdarstellung Lösungsansätze bei Problemen bezüglich der Anwendung der Standards aufzunehmen (Kategorie 314):

Erweiterung um Lösungsansätze
„Lösungsoptionen der Umsetzung könnten aufgeführt werden“
„schließlich sind Lösungsvorschläge von Interesse, wie damit umgegangen wird, falls Standards möglicherweise im Widerspruch zueinander stehen.“

Kategorie 314 listet die übrigen Beiträge auf.

Eine weitere Kategorie (Kategorie 315; 18 Nennungen) im Bereich „inhaltliche Verbesserungsvorschläge“ umfasst Beiträge, die Erläuterungen zur Anwendung der Standards in verschiedenen Kontexten vorsehen:

Erläuterungen zur Anwendung der Standards in verschiedenen Kontexten
„explizite Ausdehnung auf weitere Anwendungsfelder“
„an den spezifischen Anforderungen einzelner Fachbereiche im Bezug auf die Evaluationsstandards sollte weiter gearbeitet werden, um daraus ggf. die Bedarfe an Revision abzuleiten“
„Besonderheiten aus den verschiedenen Evaluationsschwerpunkten sollten stärker eingebunden werden, z.B. Was ist das spezifische der Evaluation in den Handlungsfeldern der Sozialen Arbeit.“

Mit 17 expliziten Nennungen erhält diese Thematik eine gehobene Bedeutung, zumal ergänzend zu diesen expliziten Nennungen, im Rahmen weiterer Beiträge, immer wieder die (spezifische) Verwendung der Standards in verschiedenen Kontexten angeschnitten wird. Demzufolge empfiehlt sich, den Schwerpunkt „feldspezifische Standards“ im Vorfeld der Revision zur Diskussion zu stellen. Ferner wird im Kapitel „Issues“ weitergehend auf diese Themenstellung eingegangen.

Die Beiträge der folgenden Kategorie 316 (15 Nennungen) nehmen ebenfalls einen beachtlichen Raum innerhalb der Antwortgesamtheit dieses Kategoriensystems ein. Auch in diesem Fall existieren neben den expliziten Nennungen viele latente Hinweise innerhalb der offenen Antworten der Befragungsteilnehmenden. Die Beiträge dieser Kategorie drücken den Wunsch aus, eine Erweiterung zur Anwendung der Standards im Rahmen bestimmter Evaluationsformen zu einem Schwerpunkt der Revision zu machen. Genauso, wie die Beiträge in Kategorie 315

die Aufarbeitung diverser Evaluationsfelder nahe legen, werden in den Beiträgen von Kategorie 316 Ausführungen zur Verwendung der Standards in unterschiedlichen Evaluationsformen gefordert: „Einzelne Evaluierungsbereiche und/oder Formen sollten beispielhaft aufgearbeitet werden“. Im Falle der Evaluationsformen erscheint dies auf den ersten Blick allerdings etwas einfacher als bei der ambivalenten Konstellation „feldspezifische Standards versus feldunspezifische Standards“. Zu folgenden Evaluationsformen (in den meisten Fällen handelt es sich um „Evaluations-Dimensionen“) wird eine Erweiterung zur Anwendung gefordert:

formative / summative Evaluation	Fremd- / Selbstevaluierung
„Generell wäre auch denkbar, nicht nur auf die summative Evaluation, sondern auch stärker auf die formative Evaluation einzugehen: damit deutlich wird, dass auch z.B. Selbstevaluationen und Evaluationen, bei denen Reflexion und Lernen stärker im Vordergrund stehen, Evaluationen sind - selbst wenn für sie dann teilweise andere Standards gelten.“	
„welche Standards besonders für eine summative oder formative Evaluierung relevant sind - welche für eine fremd oder Selbstevaluierung.“	
„In dem anzustrebenden Werk (nach Art des Joint Committee) sollten auch die Richtlinien (oder wie immer sie heißen werden) für Selbstevaluation enthalten sein.“	

Metaevaluation
„Kenntnisse über Relevanz und Nützlichkeit der Standards sind bei Evaluatoren und auch auf Auftraggeberseite vor allem am Anfang der beruflichen Karriere zu wenig verbreitet. Deshalb müsste der Stellenwert insbesondere im Hinblick auf Meta-Evaluationen im einführenden Text besser veranschaulicht werden.“

interne / externe Evaluation
„Die Spezifika der Beachtung der Standards bei interner versus externer Evaluation sind in den verschiedenartigen Anwendungsfeldern der zu evaluierenden Politikbereiche deutlicher herauszustellen. Eine getrennte Entwicklung von Standards für interne/externe Evaluation ist nicht sinnvoll, da sie eher zur Verwirrung beiträgt.“

Die Beiträge in Kategorie 317 (9 Nennungen) befassen sich mit der Interaktion zwischen Evaluierenden, Evaluierten und Auftraggebenden im Rahmen eines Evaluationsprojekts. Dabei steht die Kommunikation zwischen diesen Beteiligtegruppen im Vordergrund (*Erweiterungen ...* „hinsichtlich der Kommunikation EvaluatorInnen-AuftraggeberInnen-NutzerInnen“). In diesem Zusammenhang stellt die Kenntnis der verschiedenen Perspektiven der Beteiligten und die Fähigkeit sich diese zu vergegenwärtigen einen entscheidenden Faktor dar:

Erweiterungen zur Kommunikation / Interaktion innerhalb eines Evaluationsprojekts
„Einbezug der unterschiedlichen Perspektiven von verschiedenen, an einem

Evaluationsprozess beteiligten Personengruppen“
<i>Erläuterungen über ... „die Rollen, Funktionen und Verantwortlichkeiten der verschiedenen Beteiligten und Betroffenen (z.T. identisch, z.T. nicht, je nach der Ebene (Mitarbeiter Ebene, Zielgruppenebene, usw.)) in verschiedenen Evaluationen“</i>

Vor allem der Dyade Evaluierende-Auftraggebende kommt in den Beiträgen der Kategorie 317 eine wichtige Rolle zu:

Interaktion Evaluierende-Auftraggebende
„Es müsste Standard sein, dass Evaluatoren mit den Auftraggebern ausführlich und explizit Maßstäbe entwickeln, die der Bewertung dann zugrunde liegen können. Sonst können die Auftraggeber hinterher nach Belieben jedes Ergebnis schlecht reden.“
„Hier könnten dann evtl. auch Optionen aufgezeigt werden, wie sich EvaluatorInnen dann z.B. gegenüber dem Auftraggeber verhalten könnten?“

Kategorie 318 (sechs Nennungen) vereint Beiträge, die für sich die erweiterte Darstellung von bestimmten Zusammenhängen zum Verständnis von Evaluationen und deren Wirkungsweise aussprechen.

„Erweiterungen könnte ich mir evtl. hinsichtlich der Seite der Wirkung von Evaluationen vorstellen“

„Die Standards sollten zumindest auch Standardkriterien der Beurteilung (vgl. DAC) beinhalten, einen Minimalkatalog und eine erweiterte Liste.“

„Es müssten auch inhaltliche Punkte hinzu, die bei jedem Evaluationsprozess zu untersuchen wären, wie z.B. die zwischenmenschlichen Beziehungen, die Arbeitsprozesse, Rahmenbedingungen, die Kundenorientierung, um mal Beispiele zu nennen“

Die Beiträge dieser Kategorie drücken grundsätzlich das Bedürfnis aus, Informationen für ein tiefer gehendes Verständnis des Kosmos Standards / Evaluation zu erhalten. Im Rahmen einer Revision der Standards sollten demzufolge Überlegungen angestellt werden, derartige Informationen / Hinweise für eine neue Publikation aufzubereiten.

Die Kategorien mit inhaltlichen Verbesserungsvorschlägen sind hiermit vorgestellt worden.

In der folgenden Metakategorie 320 (zwölf Nennungen) werden Beiträge zusammengelegt, die eine **stärkere Orientierung der Standards an der Praxis** fordern. Eine konkrete Anwendungsorientierung sei momentan wichtiger als eine theoretisch wissenschaftliche Herangehensweise an diese Thematik.

stärkere Orientierung der Standards an der Praxis
„Praxisnäher, weniger wissenschaftlich“
„Anwendungsorientierung statt einseitiger Grundsatzdiskussion“
„mir fällt im Moment kein Punkt ein, an dem ich die Standards selbst inhaltlich

verändern würde, wichtiger ist mir eine Ergänzung für Praktiker. Dies aus zwei Gründen: Ich bin der Überzeugung, dass Evaluationen zukünftig in vielen Organisationen alltäglicher werden und damit auch für nicht - Expert/innen Evaluationswissen praxisorientiert aufbereitet werden sollte. Außerdem finde ich, dass eine stärkere Konkretisierung in Richtung Handreichung zu einem höheren Bekanntheitsgrad der Standards führen und ihre Akzeptanz größer würde."

Metakategorie 320 besitzt keine weiteren Untergliederungen.

Der im folgenden vorzustellenden Metakategorie 330 kommt im Zusammenhang mit der thematischen Ausrichtung einer Revision der Standards eine bedeutsame Rolle zu. Sie sammelt **Kommentare und Anmerkungen zu den Standards (und / oder den Begleittexten, Erläuterungen)**. Dabei werden in den Untergliederungen 333-339 ganz konkrete Kommentare zu spezifischen Bestandteilen der DeGEval Standards-Broschüre aufgelistet, etwa zu einem Einzelstandard oder einer bestimmten Standardgruppe o.Ä..

In Kategorie 330 (14 Nennungen) sind all die Kommentare eingefügt, die grundsätzliche bzw. in ihrer Bedeutung den ganzen Standardkosmos betreffende Äußerungen sammelt. Insofern diese Kommentare eindeutig positiver oder negativer Art sind, werden sie gesondert in Kategorie 331 (positive Kommentare; 13 Nennungen) bzw. Kategorie 332 (negative Kommentare; 18 Nennungen) abgelegt. Die „neutralen Kommentare“ in Kategorie 330 betreffen außer den bereits erwähnten grundsätzlichen Kommentaren auch Aspekte, die in dieser Form nur vereinzelt auftreten und hier nicht alle aufgeführt werden können. Exemplarisch werden einige Aspekte / Zitate vorgestellt:

grundsätzlicher Kommentar

„Ein universales Evaluationskonzept, das der kleinste gemeinsame inhaltliche Nenner aus den Praxiserfahrungen ist.“

Konflikt „verbindliche versus unverbindliche Standards“ bzw. „spezifische versus unspezifische Standards“ (siehe auch Kapitel „Issues“)

„Wichtig wäre m.E. bei der Formulierung klar zu unterscheiden zwischen einer eher abstrakten auf Dauer angelegten Metaebene und einer konkreteren Ebene, die auf Handlungsoptionen für die jeweiligen Ebenen/Verfahren/ Zielgruppen ausgerichtet ist.“

Anwendungskombination: DeGEval-Standards mit Joint Committee Handbuch

„Für meine Arbeit reichen die ausführlichen Joint Committee Standards, ergänzt um die bereits vorliegende Kurzfassung der DeGEval Standards völlig aus. Ich rate davon ab, hier weitere Ressourcen zu investieren, wie es offenbar diskutiert wird. Wir haben in Deutschland im Bereich Evaluation genug andere Defizite, die unsere Anstrengungen dringender benötigen.“

sich überschneidende Standards

„Problematisch finde ich die Standards, die sich mit anderen Teilweise überschneiden. Aber es soll ja nur ein Standardisierungsversuch sein, daher ist dieses Defizit verschmerzbar.“

13 Nennungen umfasst die Kategorie 331 (positive Kommentare / Anmerkungen), und unterstreicht damit die Wertschätzung, die den Standards entgegengebracht wird. Die in diesen Beiträgen angesprochenen Themenkomplexe sollen exemplarisch durch die folgenden Teilnehmendenzitate repräsentiert werden.

Standards bieten Orientierung

„Evaluation ist mittlerweile ein Modewort für jede Art von Auswertung. Es ist erfreulich, wenn Standards existieren, an denen man sich orientieren kann.“

Offenheit der Standards

„Der große Vorteil der Standards ist, dass sie nicht wie die meisten Publikationen über Evaluation einen bloß kleinen Bereich möglicher Evaluationen im Auge haben und Vorstellungen entwickeln, die für andere Bereiche und aus anderen Perspektiven absurd anmuten. Was in den Standards formuliert ist, ist weitestgehend für alle Bereiche möglicher Evaluation anwendbar - und damit als Standard für eine Gesellschaft, die unterschiedlichste Bereiche umfasst angemessen. Auch für Vertreter unterschiedlicher epistemologischer und methodologischer Positionen sollten sich keine großen Probleme mit den Standards ergeben, da darauf geachtet wurde, dass der kleinste gemeinsame Nenner möglichst eindeutig und präzise formuliert wurde.“

Standards als praxisorientierte Leitlinien

„Ich halte gerade die Praxisorientiertheit der Standards für ihre große Stärke, Evaluatoren / Evaluatorinnen brauchen für ihre Arbeit Leitlinien (ähnliche derer für Psychologen/Psychotherapeuten), an denen sie ihren eigenen Arbeitsansatz überprüfen und messen können. Auch die Auftraggeber sind nicht an langen theoretischen Anhandlungen zur Methodenwahl interessiert, sondern wollen für ihre jeweilige Praxis handhabbare Ergebnisse.“

Standards als Legitimation / Absicherung

„Transparenz der gängigen Standards ist für mich wichtig, um vergleichen zu können ob das, was ich plane und durchführe diesen Standards entspricht oder ggf. davon abweicht. Dies hat einen Vorteil im Hinblick auf die Argumentation (quasi legitimatorischen Charakter) und macht die Befunde der Evaluation weniger angreifbar. Gleichzeitig macht es mich persönlich sicherer in meiner Arbeit, die ich als sehr anspruchsvoll und kreativ empfinde.“

Kategorie 332 (17 Nennungen) beinhaltet die durch die Teilnehmenden geäußerten negativen Äußerungen zu den Standards. Auch hier bietet es sich an, zu den

verschiedenen angesprochenen Aspekten Zitate zur Verdeutlichung anzufügen. Da Kritik grundsätzlich konstruktiv verwertet werden sollte, erscheint es ratsam, die aufgeführten kritisierten Themenkomplexe weitestgehend vollständig darzustellen (auch bei nur einmaliger Nennung).

Standards sind nicht Öffentlichkeitswirksam

„Als schwierig in der Anwendung zeigt sich allerdings dass sich "evaluationsferne" Partner beim Anblick der "Vielzahl von Standards" einem bedrückenden und nicht einlösbaren Anspruch ausgesetzt fühlen. Daher nutze ich die Broschüre für meine Vorbereitungen und wähle die Standards genau aus, die ich berücksichtigt wissen möchte.“

Konflikt: Maximalstandards versus Minimalstandards

„Ich halte es nicht für sinnvoll, dass das Prinzip der Maximalstandards durch die Einführung unterschiedlicher Grade der Verbindlichkeit aufgeweicht wird.“

Standards haben keinen praktischen und inhaltlichen Bezug

„These: Ich glaube, dass keine einzige Arbeitsgruppe die Evaluationsstandards in irgendeinem praktischen Projekt direkt und umfassend angewandt haben. Daher ist es "nice to lesen", aber hat kaum praktischen Bezug und vor allen Dingen inhaltlichen Bezug.“

Funktion der Qualitätssicherung wird nicht erfüllt

„Aus der eigenen Arbeit lässt sich leider sagen, dass die Standards ihre Funktion der Qualitätssicherung noch nicht erfüllen, sondern lediglich einen - noch recht unbekanntem - Orientierungsrahmen darstellen. Dies ist m.E. unbefriedigend!“

Abgrenzung zu den SEVAL-Standards nicht sinnvoll

„Für mich persönlich sind die zu den SEVAL-Standards vorgenommenen Änderungen wenig überzeugend. Hier würde ich am ehesten Handlungsbedarf sehen.“

Übertragbarkeit der Standards in Transitionsgesellschaften ist fraglich

„Wenn das auf Russisch vorliegt, soll das heißen, dass die Standards auch in ehemals Ostblockländern Anwendung finden sollen. [...] Halte die Anwendung von diesen Standards in Transitionsgesellschaften, die völlig anders funktionieren, für nicht unproblematisch. Das führt zu Friktionen.“

Standards haben keinen allgemeinen Gültigkeitsanspruch

„Ich halte die Vielfalt der Evaluationen durch die gegenwärtigen Standards nicht angemessen berücksichtigt. Die ursprünglichen Standards stammen aus dem

Bildungsbereich und betrafen dort ausschließlich die Makrodidaktik. Dies scheint mir eine angemessene Einschränkung der Reichweite der Evaluationsstandards zu sein, die nicht einen Allgemeingültigkeitsanspruch an alle Evaluationen haben.“

Die folgenden Untergliederungen (Kategorien 333-339) von Kategorie 330 umfassen, wie bereits erwähnt, Kommentare und Anmerkungen, die ganz gezielt bestimmte Bestandteile der Standards bzw. der Standardsbroschüre betreffen. Da derartige Informationen konkrete Hinweise für Inhalte einer Revision der Standards darstellen könnten, wurde unabhängig von der Anzahl der Nennungen zu jedem Bestandteil der DeGEval-Broschüre (Einleitung; Erläuterungen, Standardgruppen, Einzelstandards, funktionales Inhaltsverzeichnis, sonstige Bestandteile) eine Kategorie eingerichtet. Abgesehen davon werden in Kategorie 333 (zwei Nennungen) Beiträge gesammelt, die explizit die Aufnahme eines neuen Standards fordern. Folgende Vorschläge für neue Standards wurden gemacht:

1. „Den K1 der US-Standards zur Unterstützung der Dienstleistungsorientierung fände ich nach wie vor einen wichtigen Standard, den man aufnehmen sollte.“
2. „Die Regel, dass man die Nichtanwendung von Standards deklariert, sollte zu einem eigenen Standard gemacht werden. Sie geht im Text unter. Damit können die Standards auch leicht in legitimatorischer Absicht missbraucht werden.“

Kategorie 334 (zwei Nennungen) beinhaltet Beiträge zur Einleitung in der DeGEval-Standards Broschüre. Die zwei Beiträge nehmen gegensätzliche Positionen ein. Zum einen wird erwähnt, dass die Einleitung wichtige Begrifflichkeiten definiert und somit kommunikationserleichternd wirksam ist („die Einleitung der Broschüre definiert viele wichtige Begriffe, was Kommunikation erleichtert.“). Zum anderen wird deutlich gemacht, dass diverse Aspekte in der Einleitung noch nicht angemessen berücksichtigt sind:

„Kenntnisse über Relevanz und Nützlichkeit der Standards sind bei Evaluatoren und auch auf Auftraggeberseite vor allem am Anfang der beruflichen Karriere zu wenig verbreitet. Deshalb müsste der Stellenwert insbesondere im Hinblick auf Meta-Evaluationen im einführenden Text besser veranschaulicht werden.“

Kommentare / Anmerkungen zu den Erläuterungen der Standards für Evaluation sind in Kategorie 335 (sechs Nennungen) aufgeführt. Neben der Forderung, die Standards sollten noch umfassender erläutert werden (genannte Schwerpunkte zur Ergänzung: Sondersituationen von Evaluationen, Instrumente zur Vorgehensweise / Umsetzungsvorschläge, Bezugnahme zur deutschen Evaluationspraxis) wird auch verlangt, diese genauer bzw. schärfer zu formulieren:

„In den Erläuterungen werden Begrifflichkeiten immer noch unscharf formuliert und teilweise nicht erläutert (z.B. wird Fairness durch Fairness erläutert).“

„In den Erläuterungen wird der Stellenwert von Methoden der empirischen Sozialforschung nicht ganz klar. Man kann aus dem Text schließen, dass diese eine Evaluation erst ausmachen, man kann den Text aber auch anders interpretieren. In der Definition von Evaluation wäre noch ein wenig Konkretisierungsarbeit zu leisten“

Kategorie 336 (fünf Nennungen) listet die Kommentare / Antworten zu den verschiedenen Standardgruppen auf. Drei Nennungen beziehen sich dabei auf die Genauigkeitsstandards und eine Nennung auf die Nützlichkeitsstandards:

Nützlichkeitsstandards

„Evaluation wird hier und anderswo nur nützlich, wenn sie von Entscheidern,

Führungskräfte als solche erkannt wird und die Evalinformationen in ihre Entscheidungsprozesse einbezogen werden. Dabei konkurrieren sie mit vielen anderen Informationsquellen. Ich würde mir also wünschen, bei den Nützlichkeitsstandards darauf einzugehen, wie es gelingen kann, an die Ebene der Entscheider anzudocken.“

Genauigkeitsstandards	
	„Die Standards, die sich mit nicht methodologischen Kriterien befassen, müssten mehr bekannt gemacht werden. Will man Meta Evaluationen durchführen und ist auf Veröffentlichte berichte etc. angewiesen so findet man in der Regel hauptsächlich die Genauigkeitsstandards bedient (was der Veröffentlichungspraxis wissenschaftlicher Journals entspricht).“
	„Es fehlen im Teil G Kriterien zur Nachvollziehbarkeit von Urteilen (z.B. Dokumentation und Nachvollziehbarkeit von Relevanz- und von Gütekriterien). Teilweise wird es nicht möglich sein, Daten zu gewinnen, sodass die Aussagen in G 5 sich relativieren“
	„Kritische Überprüfung der Standards der Genauigkeit, besonders vor dem Hintergrund von nutzenorientierten Selbstevaluationen sowie der ebenfalls als notwendig erkannten Verbindung von Evaluation und der Steuerung einer solchen Intervention mittels organisationsentwickelnden Maßnahmen.“

Kategorie 337 nimmt innerhalb der 330er Kategorien „Kommentare / Anmerkungen zu den Standards“ eine Sonderstellung ein. In ihr befinden sich Äußerungen zu einzelnen DeGEval-Standards. In drei Untergliederungen können die Beiträge noch einmal aufgeteilt werden, und zwar nach Beiträgen, die sich explizit zum Name eines Standards (337A, 2 Nennungen), zum Standardtext (337B, 7 Nennungen) oder zu den Standarderläuterungen (337B, 1 Nennung) äußern. Alle übrigen Beiträge zu Einzelstandards sind in Kategorie 337 aufgenommen.

Zu folgenden Standards wurden explizit Kommentare abgegeben (Kategorie 337):

N4	„N 4 ist für mich schwer zu händeln, da ich klar den Auftraggeber berücksichtige, aber Schwierigkeiten bei anderen Adressaten sehe.“
F4	„Hinsichtlich der erst kürzlich im AK "Entwicklungspolitik" diskutierten Anwendung bei Selbstevaluationen würde ich aber nicht vom zumindest auch dann angestrebten Ideal der Unparteilichkeit (s. F4) abgehen wollen. Man sollte doch selbst in der Lage sein, seine eigenen Schwächen und Fehler zu erkennen und dafür einzustehen!“
	„DE steht im Folgenden für Standards der DeGEval; JC für Standards des Joint Committee; SE für die Standards der SEVAL. Ich halte es für notwendig, diese drei Veränderungen der JC-Standards durch die DeGEval rückgängig zu machen: Die in den Standards enthaltene Zusammenfassung von <i>JC-K7: Deklaration von Interessenkonflikten</i> und <i>JC-G11: Unparteiliche Berichterstattung</i> zu <i>DE-F4: Unparteiliche Durchführung und Berichterstattung</i> führt dazu, dass der wichtige Aspekt der Thematisierung von Interessenkonflikten verloren geht. [...]“
F2	(DE steht im Folgenden für Standards der DeGEval; ... <i>siehe oben</i>) „Die ebenfalls in den DE-Standards vorgenommene Zusammenlegung von JC-

	K3: Schutz individueller Menschenrechte und JC-K4: Human gestaltete Interaktion zu DE-F2: Schutz individueller Rechte beinhaltet die Gefahr, dass unklar bleibt, dass es zwei Ebenen gibt, auf die sich dieser Standard bezieht: einerseits die Ziele, Erhebungsverfahren und Fragestellungen der Evaluation und andererseits die Interaktionen der Evaluatoren mit den Beteiligten & Betroffenen. [...]
G9	(DE steht im Folgenden für Standards der DeGEval; ... <i>siehe oben</i>) „In <i>JC-G12</i> und <i>SE-G10: Meta-Evaluation</i> ist enthalten, dass eine Meta-Evaluation auf den Evaluationsprozess anleitende formative Weise vorgenommen wird, während <i>DE-G9</i> auf diesen Aspekt verzichtet. [Ansonsten sehe ich keine grundlegenden Mängel an den Standards.]

Zwei Beiträge, die in Kategorie 337A abgelegt sind, beziehen sich auf bestimmte Standardnamen, die revidiert werden könnten. Es handelt sich dabei um Standard *D2: Diplomatisches Vorgehen* sowie Standard *F4: Unparteiische Durchführung und Berichterstattung*:

D2	„D 2 "Diplomatisches Vorgehen". Der Begriff gefällt mir nicht, da er für mich immer wieder mangelnde Offenheit bedeutet.“
F4	„So ist nach meiner Meinung die Formulierung des Standards F4 „Unparteiische Durchführung und Berichterstattung“ revisionsbedürftig. „Unparteilichkeit“ ist eine Idealvorstellung, die sich nicht mit der Praxis deckt, denn die Evaluationen / Evaluationsergebnisse orientieren sich immer stärker an den Interessen und Werten des Auftraggebers als an politischen oder wissenschaftlichen Objektivitätsvorstellungen“

Zu sechs DeGEval-Standards sind in Kategorie 337B Kommentare / Anmerkungen hinsichtlich der Ausformulierung des jeweiligen Standardtextes gemacht worden. Dabei spielen in erster Linie formale Aspekte, wie z.B. der Umgang mit Passivkonstruktion bei der Formulierung des Standardtextes eine Rolle. Allerdings haben die Wirkungen derartig formulierter Standards mitunter nicht gewollte inhaltliche Auswirkungen hinsichtlich Verständnis und -in einem zweiten Schritt- auch der Art der Verwendung / Beachtung des Standards. Die Autoren machen zum Teil alternative Formulierungsvorschläge.

Der Text folgender Standards gelangt durch die Beiträge in den Fokus einer möglichen Überarbeitung:

N3	„Zu den eher formalen Vorschlägen gehört der Umgang mit Passivkonstruktionen. Sie sollten m. E. in den Standards so wenig wie möglich stehen. Zu ändern ist z.B.: Statt bisher: <i>N3 Glaubwürdigkeit und Kompetenz des Evaluators / der Evaluatorin</i> : Wer Evaluationen durchführt, soll persönlich glaubwürdig sowie methodisch und fachlich kompetent sein, damit bei den Evaluationsergebnissen ein Höchstmaß an Glaubwürdigkeit und Kompetenz erreicht wird. Zu: <i>N3 Glaubwürdigkeit und Kompetenz des Evaluators / der Evaluatorin</i> : Wer Evaluationen durchführt, soll persönlich glaubwürdig sowie methodisch und fachlich kompetent sein, damit die Evaluationsergebnisse durch ein Höchstmaß an Glaubwürdigkeit und Kompetenz überzeugen.“
-----------	--

N4	„Manchmal zu komplexe Sätze. Müssen alle Aussagen in einen Satz stehen? Statt <i>N4 Auswahl und Umfang der Informationen</i> : Auswahl und Umfang der erfassten Informationen sollen die Behandlung der zu untersuchenden Fragestellungen zum Evaluationsgegenstand ermöglichen und gleichzeitig den Informationsbedarf des Auftraggebers und anderer Adressaten und Adressatinnen berücksichtigen. Neue Fassung: <i>N4 Auswahl und Umfang der Informationen</i> : Das Evaluationsteam soll Informationen soweit auswählen und erfassen, dass es die Fragen zum Evaluationsgegenstand angemessen behandeln kann. Dabei soll es den Informationsbedarf des Auftraggebers und den anderer Adressaten und Adressatinnen berücksichtigen.“
-----------	--

D1	„In einigen Standards, wie in D1, erklingt eine sehr formale und technische Sprache. Ist hier nicht ein einfacherer Ausdruck möglich? Was unter einem „Evaluationsgegenstand“ zu verstehen ist, erläutert bereits G1. Obwohl in D1 der Gebrauch des Wortes „Evaluationsgegenstand“ sicher logisch richtig ist, klingt dies aber im Zusammenhang mit „Belastung“ wie ein technischer Test, etwa wie die Überprüfung, bis zu welchem Gewicht sich noch Fahrzeuge auf einer Brücke bewegen können. Deshalb würde ich vorschlagen, D1 zu kürzen. Die Erläuterung zu D1 ist dann m.E. zur näheren Bestimmung ausreichend. Statt: <i>D1 Angemessene Verfahren</i> : Evaluationsverfahren, einschließlich der Verfahren zur Beschaffung notwendiger Informationen, sollen so gewählt werden, dass Belastungen des Evaluationsgegenstandes bzw. der Beteiligten und Betroffenen in einem angemessenen Verhältnis zum erwarteten Nutzen der Evaluation stehen. Neue Version + Erläuterung wie bisher <i>D1 Angemessene Verfahren</i> : Evaluationsverfahren und Maßnahmen der Informationsbeschaffung können Personen und Abläufe stören. Die Belastungen, die durch diese Verfahren und Maßnahmen entstehen, sollen in einem angemessenen Verhältnis zum erwarteten Nutzen der Evaluation stehen.“
-----------	--

F4	„ <i>F4 Unparteiische Durchführung und Berichterstattung</i> ist zu hart und wissenschaftstheoretisch zu einseitig formuliert. Damit werden parteiische Evaluationsansätze - auch wenn ich persönlich diese nicht favorisiere - grundsätzlich ausgeschlossen. Deklaration von Interessenskonflikten wäre sachangemessener.“
-----------	---

	<p>„Etwa Formulierung von F4 so, dass die Zielsetzung des Evaluationsteams deutlich wird. Statt: <i>F4 Unparteiische Durchführung und Berichterstattung</i>: Die Evaluation soll unterschiedliche Sichtweisen von Beteiligten und Betroffenen auf Gegenstand und Ergebnisse der Evaluation in Rechnung stellen. Berichte sollen ebenso wie der gesamte Evaluationsprozess die unparteiische Position des Evaluationsteams erkennen lassen. Bewertungen sollen fair und möglichst frei von persönlichen Gefühlen getroffen werden. Neue Formulierung: <i>F4 Unparteiische Durchführung und Berichterstattung</i>: Die Personen, die von der Evaluation betroffen sind, haben unterschiedliche Sichtweisen zum Gegenstand und den Ergebnissen der Evaluation. Die Evaluation soll diese unterschiedlichen Sichtweisen zeigen. Der gesamte Evaluationsprozess soll verdeutlichen, dass das Evaluationsteam eine moderierende, unparteiische Position anstrebt. Bewertungen sollen fair und möglichst frei von persönlichen Gefühlen getroffen werden.“</p>
--	---

F5	F 5 Die Offenlegung der Ergebnisse nicht "so weit wie möglich" sondern ohne diesen Zusatz.
-----------	--

G3	<p>„Aber auch aus eher inhaltlichen Erwägungen sollten einige Passivsätze nicht mehr im Text stehen. Statt: <i>G3 Beschreibung von Zwecken und Vorgehen</i>: Gegenstand, Zwecke, Fragestellungen und Vorgehen der Evaluation, einschließlich der angewandten Methoden, sollen genau dokumentiert und beschrieben werden, sodass sie identifiziert und eingeschätzt werden können. Genauere Bestimmung des handelnden Subjekts / der verantwortlichen Subjekte, beispielsweise: <i>G3 Beschreibung von Zwecken und Vorgehen</i>: Das Evaluationsteam oder die für Dokumentationsarbeiten verantwortlichen Personen sollen Gegenstand, Zwecke, Fragestellungen und Vorgehen der Evaluation, einschließlich der angewandten Methoden, genau dokumentieren und beschreiben, sodass jede unmittelbar an der Evaluation beteiligte Person, aber auch die Öffentlichkeit den Verlauf der Evaluation identifizieren und einschätzen kann.“</p>
-----------	--

Auch die Erläuterungen zu den Standards sind kommentiert worden. Ein Beitrag in Kategorie 337C nimmt zu den Erläuterungen des Standards F4 Stellung:

„Die Evaluation übernimmt, so weit ich das einschätzen kann, fast nie nur die spezifische Sichtweise einer Partei, aber sie legt den Öfteren den Schwerpunkt auf die Sichtweise einer Partei: den Auftraggeber. Statt: „Die Evaluation soll darauf verzichten, lediglich eine spezifische Sichtweise zu übernehmen. Vielmehr soll sie eine faire Behandlung aller relevanten Interessen anstreben ...“ Vorschlag: „Die Evaluation soll darauf verzichten, die Sichtweise einer Partei zu bevorzugen. Vielmehr soll sie eine faire Behandlung aller relevanten Interessen anstreben“

Kategorie 338 (zwei Nennungen) umfasst zwei Beiträge, die das funktionale Inhaltsverzeichnis der DeGEval-Broschüre kommentieren. Die Wichtigkeit dieses Instruments wird durch die Kommentare unterstrichen.

„wichtig ist für mich auch das funktionale Inhaltsverzeichnis: worauf muss ich in welcher Evaluationsphase achten, um einen erfolgreichen Verlauf zu ermöglichen“

„Organisation der Standards nach dem Ablauf einer Evaluation gegliedert, dann sind sie besser einsetzbar. Widersprüche zwischen den Standards werden dann auch deutlicher (dass man nicht jedem Standard in gleicher Weise gerecht werden kann)“

Schließlich bietet Kategorie 339 (eine Nennung) Platz für Kommentare / Anmerkungen zu sonstigen Einzelbestandteilen. Der Beitrag befasst sich mit der in der Standards-Broschüre definierten Begrifflichkeit „Evaluationsforschung“. Um Irritationen im Vergleich zum internationalen Sprachgebrauch dieses Begriffes zu vermeiden, wird eine Umschreibung empfohlen:

„Der Begriff Evaluationsforschung wird in den Standards in einer sehr spezifischen Weise definiert, die nicht dem international üblichen Sprachgebrauch entspricht. Dies mag gute Gründe haben, führt an dieser Stelle jedoch eher zu Irritationen. Man kann diesen Begriff im letzten Standard auch einfach umschreiben. Das fände ich angemessener, da mit den Standards doch eher die Fachdebatte aufgenommen werden sollte, damit sie breit akzeptiert werden, als dass relativ spezifische Positionen formuliert werden sollten, die mit gewisser Wahrscheinlichkeit Gegenpositionierungen provozieren.“

Mit Metakategorie 340 existiert eine weitere bedeutsame Kategorie. Sie umfasst alle Beiträge, die den angedachten **Revisionsprozess kommentieren**. Die zwei Untergliederungen ermöglichen eine Differenzierung der Kommentare nach gesonderten Schwerpunkten. Dabei sind in Kategorie 341 (14 Nennungen) konkrete Vorschläge der Befragungsteilnehmenden zur Umsetzung und Durchführung der Revision aufgelistet. Die Befragungsteilnehmenden haben sich hierzu sehr ausführlich geäußert. Die folgenden Beiträge spiegeln die meistgenannten Vorschläge exemplarisch wieder:

strukturiertes Revisionsverfahren
„Der Revisionsprozess muss zunächst diese a) inhaltlichen Evaluationsthemen sammeln (schriftliche Einreichung von Evaluationskriterien) b) Kommissionkreis wird die Themen Gruppieren, Clustern und Doubletten rausnehmen c) Das so gefundene Konzept mit Oberthemen wird dann mit jedem Kriterium/Oberthema in die Arbeitsgruppen gegeben, um dann von der Basis (mit Externen) die Details ausarbeiten zu lassen.“
„Ich weiß nicht, ob es sie interessiert: in RéseauEval hier ist es ein Prozess in 5 Schritten: 1) ein Comité redigiert texte und organisiert einen Austausch via E-Mail. 2) es integriert die Bemerkungen weitgehend 3) der Vorstand nimmt Stellung zum verbesserten Text 4) die Hauptversammlung validiert die Ergebnisse 5) das Ergebnis erscheint im Internet.“

Beteiligtenkreis ausweiten
„Der Revisionsprozess sollte auf keinen Fall auf eine kleine Kommission - und deren Nachdenken bzw. immer nur begrenzte Erfahrung - beschränkt bleiben. Eine Revision, die möglichst viele Perspektiven einbezieht, wäre hilfreich“
„Die Revision sollte kein Erguss eines Forscherteams sein, sondern ein organisierter Gruppenprozess sein, an dem alle Arbeitskreise der DeGEval ihre meistgebrauchten inhaltlichen Evaluationskriterien definieren und zusammentragen.“

Interaktives Element I: sozial interaktive Veranstaltung
„Im Revisionsprozess wäre ein sozial interaktives Element, z.B. eine große Open Space Veranstaltung ein zur Umfrage ergänzend geeignetes Instrument, um schnell

und effektiv Rückmeldung und Weiterarbeit für die Standards zu leisten“

Interaktives Element II: Expertenbefragung

„Einbezug vieler externer Experten und Praktiker. Dies könnte beispielsweise in Form einer DELPHI-Technik erfolgen.“

Interaktives Element III: Forum / Internetplattform(en) einbinden

„Man könnten zudem auch gezielt die Standards im Forum diskutieren: Pro Monat einen Standard mit Bitte um Beispiele, Erfahrungen usw. das wäre ein zu moderierender Prozess, der auch das Forum aufwerten könnte.“

„Einrichtung einer Internetplattform (Blackboard) in der jeder den Arbeitsstand der anderen Gruppen nachlesen kann. Virtuelle Zusammenarbeit fördert die Gruppenergebnisse.“

Kategorie 342 (sechs Nennungen) hingegen beinhaltet all die Beiträge, welche eine skeptische oder ablehnende Haltung gegenüber einer Revision der Standards ausdrücken. Im Zentrum vieler Beiträge dieser Kategorie steht die Annahme, dass die Ressourcen, die für eine Revision aufgewendet werden müssten, an anderer Stelle dringlicher benötigt werden.

„Intensive Schulung und Öffentlichkeitsarbeit, seitens der DeGEval sind, zur Erhöhung des Bekanntheitsgrades der Standards wichtiger als umfangreiche Revisionsbemühungen.“

„Im Interesse einer verbesserten Informationsbasis über Nützlichkeit und Relevanz der Standards wird eine Intensivierung der Öffentlichkeitsarbeit und Intensivierung der Thematik in den Arbeitskreisen der DeGEval für vordringlich gehalten im Vergleich zu einer umfangreichen Revisionsaktivität.“

„Andere Aufgaben scheinen mir für die DeGEval derzeit dringlicher (Anmerkung: als der Revisionsprozess)“

Kommentare zum Revisionsprozess, die nicht in einer dieser beiden Kategorien abgelegt werden konnten, werden in Kategorie 340 (15 Nennungen) aufgeführt. In diesem Fall beschreiben die Beiträge häufig, was eine Revision leisten sollte:

Was eine Revision leisten sollte / Bsp.

„Aufgabe der Revision sollte es sein, die bislang gesammelten Erfahrungen mit den Standards auszuwerten und zu prüfen, ob dies zur Weiterentwicklung der Standards Anlass gibt.“

„Die Revision sollte prüfen, ob aus Sicht der verschiedensten Evaluationsinteressierten Ergänzungen oder Änderungen notwendig sind.“

„Außerdem hat die Revision unabhängig von ihren inhaltlichen Ergebnissen auch die Funktion, die Standards in der Diskussion zu halten.“

Außerdem werden diverse Details der Revision angeschnitten und ins Bewusstsein gerufen („Ich plädiere für eine Revision bis 2008, auch um die Mitwirkungsbereitschaft zu erhöhen“, „Ich halte die Transparenz des praktizierten Verfahrens der Revision für begrüßenswert!“, ...), aber

auch Fragen aufgeworfen, die im Zusammenhang mit der Realisierung eines Revisionsverfahrens beantwortet werden müssen:

„Problematisch ist jedoch die Frage, WER die Revision durchführt bzw. WIE die Verabschiedung erfolgen soll, denn ein reiner Mehrheitsbeschluss der DeGEval-Mitglieder selbst ist notwendiger Weise kein Garant für eine hohe Qualität des Produktes, d.h., es könnte durchaus passieren, dass einige Mitglieder ein sehr anspruchsvoll formuliertes Werk nicht akzeptieren?“

Kategorie 350 (16 Nennungen), als letzte Kategorie im Kategoriensystem „Kommentare / Verbesserungsvorschläge“ sammelt alle Beiträge, die in keiner bisher vorgestellten Kategorie untergebracht werden konnten. In den meisten Fällen liegt dies daran, dass die Beiträge keinen inhaltlichen Bezug zur jeweiligen Thematik aufweisen. Dies hat unterschiedliche Gründe, z.B. fehlende Vertrautheit mit den Standards. Einige wenige Beiträge befassen sich mit der Befragung REVAL bzw. dem Befragungsdesign selbst:

„Der Fragebogen mag nicht in allen Teilen zu überzeugen. Besonders hinsichtlich der Einschränkungen auf zwei Rollen und auf ein Themengebiet schätze ich als sehr unangemessen ein, besonders wenn die DeGEval dies tut. In beiden Fällen soll ausgeschlossen werden, was ich persönlich als nützlich empfinde: Austausch zwischen Rollen und Themenfeldern.“

„Die nachfolgende Frage nach dem Zeitraum (*Anmerkung: Frage v_138175: „In welchem Zeitrahmen sollte eine Revision der Standards für Evaluation durchgeführt werden?“*) lässt sich in Unkenntnis der gewählten Option, nicht sinnvoll beantworten. Ich gehe davon aus, dass Option (4) zugrunde liegt.“

Issues

Die qualitative Datenauswertung von REVAL -Befragung zur Steuerung des Revisionsprozesses der DeGEval Standards für Evaluation- hat unterschiedliche Erwartungen an, und Vorstellungen von den DeGEval-Standards sichtbar gemacht. Diese zum Teil sehr kontroversen Ansichten bestimmen maßgeblich die Auseinandersetzung über das Ob und Wie einer Revision der Standards. Um Missverständnissen vorzubeugen empfiehlt es sich, dass die DeGEval zumindestens in einigen dieser strittigen Punkte klarstellt, was die Standards leisten können und sollen und was nicht.

Folgende Kontroversen offenbarten sich durch die Analyse der qualitativen Daten: hängen dicht zusammen ...

1. Maximal-Standards versus Minimal-Standards

Durch die Textbeiträge der Teilnehmenden wird deutlich, dass keine einvernehmliche Haltung gegenüber dem Prinzip der Maximal-Standards besteht. Es existieren zwar Beiträge, die deutlich herausstellen, dass ein Prinzip der Maximal-Standards ein angemessenes Format für Standards mit weitreichendem Gültigkeitsanspruch darstellen („Ich halte es nicht für sinnvoll, dass das Prinzip der Maximalstandards durch die Einführung unterschiedlicher Grade der Verbindlichkeit aufgeweicht wird.“), es wird allerdings auch deutlich, dass die Einführung von Minimal-Standards von einigen Befragungsteilnehmenden begrüßt werden würde („genauere Festlegung von Mindeststandards (Qualitätssicherung), ...“).

Um das Konfliktpotenzial dieser Kontroverse aufzulösen, wäre eine klare Positionierung der DeGEval hilfreich, inwiefern der Umgang mit den Standards angedacht ist:

„Wichtig wäre m.E. bei der Formulierung klar zu unterscheiden zwischen einer eher abstrakten auf Dauer angelegten Metaebene und einer konkreteren Ebene, die auf Handlungsoptionen für die jeweiligen Ebenen/Verfahren/ Zielgruppen ausgerichtet ist.“

2. knappe Standards versus detaillierte Standards

Eine weitere Kontroverse bezieht sich auf den Umfang der Standards. Hier werden sehr unterschiedliche Niveaus gefordert. Die Bandbreite geht von knappen Kurzfassungen bis hin zu ausführlich beschriebenen und erläuterten Standards. Die unterschiedlichen Anforderungen resultieren in entscheidendem Maße daraus, für welche Zwecke die Standards von wem wie verwendet werden. Es ergeben sich sehr unterschiedliche Bedarfe. In diesem Zusammenhang wird von Teilnehmenden unter anderem der Vorschlag gemacht unterschiedlich umfangreiche Formate der Standards anzubieten:

„Hilfreich wäre vielleicht eine Priorisierung der Standards. 25 sind eine ganze Menge und die muss man schon nachschlagen, weil man die nicht immer im Kopf haben kann. Vielleicht könnte man (je nach Zielgruppe) umfassendere und weniger umfassendere Standards entwickeln?“

Folgt man diesem Vorschlag gelangt man allerdings direkt in den Wirkungskreis weiterer Streitfragen (Maximal-Standards versus Minimal-Standards, feldunspezifische Standards versus feldspezifische Standards).

3. feldunspezifische Standards versus feldspezifische Standards

Besonders häufig begegnet man bei der Analyse der qualitativen Daten der Forderung, es sollen spezifische Standards für die verschiedensten Anwendungsbereiche entwickelt werden. Die Teilnehmenden, die feldunspezifische Standards präferieren, sind gegenüber denen, die feldspezifische Standards fordern, zahlenmäßig eindeutig unterlegen. In seltenen Fällen wird die feldunspezifische Ausrichtung der Standards positiv hervorgehoben:

„Ich finde die Standards im Großen und Ganzen gut, sie können als allgemeine Leitlinien gelten (als "rahmen") - natürlich sind sie recht allgemein und abstrakt formuliert, eine Anpassung an den jeweiligen Kontext muss selbst vorgenommen werden - das finde ich aber gut“

Ein Großteil der Beiträge, die sich zu dieser Streitfrage äußern, fordern ausdrücklich, dass die Standards auf unterschiedliche Anwendungsbereiche angepasst bzw. zugeschnitten werden sollten.

„... oder, das wäre allerdings wieder viel mehr Arbeit, die Entwicklung von Standards für verschiedene Evaluationsfelder. Hier sollte die DeGEval im eigenen Interesse die Vorreiterrolle spielen.“

„Die Standards müssen alltagstauglicher gestaltet werden. Dies nicht im Sinne einer Aufweichung der Standards sondern im Sinne einer Berücksichtigung der sehr unterschiedlichen Arbeitsbereiche von Evaluatoren.“

Inwiefern bei diesen Forderungen die Bequemlichkeit eine Rolle spielt, bald auf eine Anpassung der Standards je nach Kontext verzichten zu können, da spezifische Standards vorliegen, kann hier nicht beantwortet werden.

4. offene Standards versus verbindliche Standards

Die Frage, wie verbindlich die DeGEval-Standards sein sollten, hängt eng mit der Frage zusammen, wie spezifisch die Standards auf einen bestimmten Bereich zugeschnitten werden sollten. Insofern die Standards verbindlicher werden sollen, würde dies bedeuten, dass ein stärkerer Bezug zum jeweiligen Kontext hergestellt werden müsste:

„Daher sollten entweder weniger aber viel generellere Standards entwickelt werden (und diese mit zahlreichen Beispielen erläutert werden) oder der Anwendungskontext (große Evaluationen mit universitär-akademischen Allgemeingültigkeitsanspruch) detailliert erläutert werden.“

Es gilt zu berücksichtigen, dass sich offen formulierte Standards ein hohes Maß an Flexibilität z.B. hinsichtlich des Bereiches in welchem sie angewendet werden sollen, bewahren:

„Der große Vorteil der Standards ist, dass sie nicht wie die meisten Publikationen über Evaluation einen bloß kleinen Bereich möglicher Evaluationen im Auge haben und Vorstellungen entwickeln, die für andere Bereiche und aus anderen Perspektiven absurd anmuten. Was in den Standards formuliert ist, ist weitestgehend für alle Bereiche möglicher Evaluation anwendbar - und damit als Standard für eine Gesellschaft, die unterschiedlichste Bereiche umfasst angemessen. Auch für Vertreter unterschiedlicher epistemologischer und methodologischer Positionen sollten sich keine großen Probleme mit den Standards ergeben, da darauf geachtet wurde, dass der kleinste gemeinsame Nenner möglichst eindeutig und präzise formuliert wurde.“

Die Auswahl und Anpassung allgemein formulierter Standards für den Einsatz in unterschiedlichen Evaluationsprojekten, stellt allerdings einen gewissen Arbeitsaufwand dar und setzt Know-how und Erfahrung in den Bereichen Standards, Evaluation und Empirie voraus.

Ergebniszusammenfassung

Die zahlreichen und umfangreichen Textbeiträge im Rahmen der Online-Befragung zur Steuerung des Revisionsprozesses der DeGEval-Standards für Evaluation zeugen von einem sehr starken Interesse an den Standards und ihrer weiteren Entwicklung. Der Bedarf für Evaluationsstandards, der durch die Beiträge ausgedrückt wird, soll durch revidierte Standards noch besser abgedeckt werden. Dies wünschen sich zumindest zahlreiche Befragungsteilnehmende, die dieser Hoffnung durch zahlreiche Verbesserungsvorschläge und Kommentare Ausdruck verleihen. Neben Fallbeispielen und ergänzenden Hinweisen, die die Anwendung der Standards verdeutlichen, wird vor allem ein Bezug zur Verwendung der Standards in unterschiedlichen Kontexten verlangt. Abgesehen davon wird in vielen Beiträgen ein Bedürfnis nach aufbereitetem theoretischem Grundlagenwissen für die Thematik Evaluation / Standards ausgedrückt. Inwiefern derartige Informationen in einer revidierten Standardpublikation der DeGEval aufgenommen werden können, muss an anderer Stelle erörtert werden.

Die Kommentare zur Bewertung der Standards in ihrer momentanen Version sind eher allgemein formuliert worden. Nur sehr wenige Teilnehmende äußern sich ganz konkret zu bestimmten Standards deren Erläuterungen oder sonstigen Bestandteilen der DeGEval-Standards Broschüre. In erster Linie handelt es sich jedoch um positive Bewertungen der Standards mit weiterem Potenzial zur Verbesserung.