

A.1 Collaborative Filtering in Ideenwettbewerben: Evaluation zweier Skalen zur Teilnehmerbewertung

Ivo Blohm¹, Ulrich Bretschneider¹, Michael Huber¹, Jan Marco Leimeister², Helmut Krcmar¹

¹Technische Universität München, Lehrstuhl für Wirtschaftsinformatik

²Universität Kassel, Lehrstuhl für Wirtschaftsinformatik

Zusammenfassung

In diesem Beitrag werden zwei verschiedene Skalen zur Durchführung von Teilnehmerbewertungen in Ideenwettbewerben auf ihre Validität untersucht. Dafür werden diese mit einer unabhängigen, validierten Expertenbewertung verglichen. Auf Basis von Kreuztabellen wird gezeigt, dass eine einfache binäre Skala („Go“ oder „No Go“) eine höhere Übereinstimmungsvalidität besitzt als ein komplexes, mehrdimensionales Bewertungsformular. Auf Basis dieser Untersuchung werden konkrete Gestaltungsempfehlungen für die Praxis abgeleitet sowie zukünftiger Forschungsbedarf aufgezeigt.

1 Einleitung

Seit Beginn des 21. Jahrhunderts kommt es in der Innovationsentwicklung immer stärker zu einem Paradigmenwechsel. Neue Produkte werden nicht mehr allein in geschlossenen Innovationsprozessen entwickelt, sondern eine Öffnung des Innovationsprozesses für externe Einflüsse gewinnt zunehmend an Bedeutung. Gerade der aktiven Integration von Kunden wird in diesem Zusammenhang eine besonders hohe Relevanz zugesprochen (RePi06).

Ideenwettbewerbe sind in der Praxis auf Grund ihrer Einfachheit ein äußerst verbreiteter Ansatz zur aktiven Kundenintegration (Walcz07). Prominente Erfolgsbeispiele wie der IBM „Innovation Jam“, ein Ideenwettbewerb, in dem weltweit 140.000 Teilnehmer mehr als 46.000 Ideen einreichten, verdeutlichen das immen-

se Potenzial dieses Ansatzes (Kais09). Einerseits handelt es sich bei einer solchen Anzahl von Ideen um ein enormes Innovationspotential. Andererseits bedeutet dies für den Veranstalter aber auch einen hohen administrativen Aufwand, denn ein Screening der Ideen wird in der Regel durch eine kleine Gruppe von Experten vorgenommen und bindet im hohen Maße Unternehmensressourcen.

Ein verstärktes Einbinden der Teilnehmer, beispielsweise im Rahmen einer Vorauswahl der Ideen, erscheint vor diesem Hintergrund aussichtsreich. So ist es wenig verwunderlich, dass Funktionalitäten zur Durchführung von User Votings heute in vielen Internet-gestützten Ideenwettbewerben eingesetzt werden (ELBK08; LHBK09). Eine solche Teilnehmerbewertung kann für ein Unternehmen aber nur einen Mehrwert schaffen, wenn die Teilnehmer die besten Ideen effektiv herausfiltern können. Jedoch gibt es bislang nur wenig wissenschaftliche Untersuchungen zur Validität solcher Mechanismen. Die wenigen systematischen Untersuchungen kommen zu unterschiedlichen Ergebnissen. In einigen Studien konnte eine gute Übereinstimmung zwischen Teilnehmer- und Expertenbewertungen festgestellt werden (Walc07), in anderen Studien wiederum konnte dieser Zusammenhang nicht bestätigt werden (FrHi06). Der Einfluss der verwendeten Bewertungsskala auf die Güte solcher Bewertungen wurde in diesem Kontext aber noch nicht systematisch untersucht. Geeignete Bewertungsskalen sind jedoch eine wesentliche Voraussetzung zur Durchführung von validen Teilnehmerbewertungen (AvZe97; Nich97). Bei vielen Online-Rating-Mechanismen (vgl. auch WHPT09) scheinen einfachste, binäre Skalen zum Zwecke eines „Collaborative Filtering“, also eines gemeinschaftlichen Filterns von Informationen (AvZe97), gut zu funktionieren. In Ideenwettbewerben werden den Ideengebern häufig aber auch komplexere Bewertungsskalen, die mehrere Bewertungsdimensionen, wie z.B. Neuartigkeit oder Marktpotential, umfassen, zur Verfügung gestellt. Hier ist prinzipiell zu vermuten, dass diese auf Grund granularerer Bewertungsmöglichkeiten eine höhere Bewertungsgüte ermöglichen.

Ziel des folgenden Beitrages ist es, Empfehlungen für die Gestaltung von Instrumenten zur Durchführung von Teilnehmerbewertungen in Ideenwettbewerben zu

entwickeln. Dafür wurden zwei unterschiedliche Skalen zur Teilnehmerbewertung in einem Ideenwettbewerb implementiert und anschließend deren Ergebnisse mit einer Expertenbewertung verglichen. In Kapitel 2 erfolgen dafür zunächst eine Beschreibung von Ideenwettbewerben und von Konzeptualisierungen des komplexen Konstrukts Ideenqualität. In Kapitel 3 werden der durchgeführte Feldtest, die implementierten Bewertungsskalen und das Evaluationsdesign näher beschrieben. Die Ergebnisse des Feldtests werden in Kapitel 4 dargestellt und anschließend in Kapitel 5 diskutiert. Ein Ausblick rundet den Beitrag ab.

2 Theoretischer Hintergrund

2.1 Ideenwettbewerbe

Walcher (2007) definiert einen Ideenwettbewerb als eine Aufforderung eines privaten oder öffentlichen Veranstalters an die Allgemeinheit oder eine spezielle Zielgruppe, themenbezogene Beiträge innerhalb eines bestimmten Zeitraums einzureichen, die dann in der Regel von einer Jury bewertet und leistungsorientiert prämiert werden. Ideenwettbewerbe im Rahmen einer Kundenintegration in die Innovationsentwicklung haben das Ziel, Kunden zur Unterstützung der Ideengenerierung heranzuziehen. Aus der Sicht des veranstaltenden Unternehmens erweitern Ideenwettbewerbe den Ideenraum in den frühen Phasen des Innovationsprozesses. Der dabei innewohnende Wettbewerbscharakter, der durch die Ausschreibung von Prämierungen für die besten Beiträge forciert wird, soll die Kreativität und Qualität der Beiträge anregen und den Teilnehmern einen Anreiz zur Teilnahme bieten (RePi06). Ideenwettbewerbe basieren auf der Annahme, dass sich in den Kundenideen die Wünsche und Bedürfnisse der Kunden manifestieren und Unternehmen so auf effiziente Weise an die Bedürfnisinformationen ihrer Kunden gelangen können (BrLK07; BrLK09). Je nach Ausgestaltungsform des Ideenwettbewerbes können die Kundenideen auch Lösungsinformationen enthalten, also konkrete Informationen darüber, wie die vorgeschlagenen Ideen umgesetzt werden könnten.

2.2 Dimensionen von Ideenqualität

Die Bewertung von Innovationsideen in den frühen Phasen des Innovationsprozesses ist eng verbunden mit der Bewertung der diesen innewohnenden Kreativität (Walcz07). Kreative Lösungen werden dabei allgemein als neuartig und angemessen definiert (Amab96; PIBD04). Neuartigkeit impliziert, dass neue Ideen für den Veranstalter unbekannt sind (MaWa94) und wird häufig auch mit Originalität gleichgesetzt. Originelle Ideen sind aber nicht nur neu, sondern auch eigentümlich, genial, fantasievoll, überraschend oder unerwartet (DHRS06). Ein weiteres Kennzeichen von Neuartigkeit ist die sog. "Paradigm Relatedness" (NaBo94; FiWS96). Dieser Paradigmenbezug bezeichnet das Ausmaß, in dem eine Idee etablierte Strukturen überwindet und beschreibt damit in einer gewissen Weise den Innovationsgrad einer Idee. Angemessenheit kann als das Ausmaß definiert werden, in dem für den Empfänger ein wesentliches Problem gelöst wird (Amab96; DHRS06). Diese Dimension wird häufig auch als Relevanz oder Wert einer Idee bezeichnet (MaWa94; KrGA04; DHRS06). In der Domäne der Innovationsentwicklung bezieht sich dies in der Regel auf das finanzielle Potential einer Idee (Roch91; CaVa99; FrHi06; Soll06).

Die Umsetzbarkeit einer Idee stellt einen weiteren wesentlichen Bestandteil von Ideenqualität dar. Im Rahmen der Neuproduktentwicklung beschreibt die Umsetzbarkeit im Wesentlichen die Einfachheit, mit der eine Idee in ein neues Produkt überführt werden kann (KrGA04; Soll06) sowie den Fit zum Organisator, dessen Strategie, Fähigkeiten und Ressourcen (Roch91). Umsetzbarkeit besitzt eine finanzielle und eine technische Dimension (Soll06). Letzte Facette von Ideenqualität ist der Konkretisierungsgrad (MaWa94; Walcz07), der im Wesentlichen den Grad einer vollständigen, detaillierten und exakten Beschreibung (DHRS06) sowie der Reife einer Idee (FrHi06) beschreibt.

Um die Qualität einer Idee messbar zu machen gibt es unterschiedliche Ansätze und Instrumente. Im Folgenden werden zwei Ansätze (eine einzelne binäre Skala versus mehrdimensionale Bewertungsskalen) zur Durchführung von Teilnehmer-

bewertungen in Ideenwettbewerben auf ihre Validität unter realen Bedingungen empirisch untersucht.

3 Untersuchungsdesign des explorativen Feldtests

3.1 Projekthintergrund: Der SAPIens Ideenwettbewerb

SAPIens ist ein Internet-basierter Ideenwettbewerb des Software-Anbieters SAP, der im Sommer 2008 mit einer Laufzeit von 14 Wochen durchgeführt wurde. Teilnahmeberechtigt waren SAP-Anwender, die aufgefordert wurden „neue Ideen rund um SAP“ zu entwickeln und auf der IT-gestützten SAPIens-Plattform einzureichen. Die Ideeneinreichung selbst erfolgte über ein Webformular, das speziell für den SAPIens Ideenwettbewerb entwickelt wurde. Nach Ende des Wettbewerbs wurden die eingereichten Ideen zur Ermittlung der Gewinner durch ein Expertengremium bewertet. Es registrierten sich 127 Teilnehmer für den Ideenwettbewerb, von denen 39 mindestens eine Idee einreichten. Insgesamt wurden 57 Ideen eingereicht. 72% der Teilnehmer waren männlich und 78% waren zwischen 20 and 25 Jahre alt. Dies deckt sich mit der Teilnehmerzusammensetzung in anderen Ideenwettbewerben (Soll06; Walc07).

3.2 Beschreibung der implementierten Bewertungsskalen

Zur Untersuchung der Validität unterschiedlicher Skalen zur Teilnehmerbewertung wurden auf der SAPIens-Online-Plattform zwei unterschiedliche Bewertungsskalen implementiert und im Feld getestet. Bei den verwendeten Skalen handelt es sich einerseits um eine binäre Skala („Go“ oder „No Go“) und andererseits um ein komplexeres, mehrdimensionales Bewertungsformular, in dem Punkte wie Originalität, Innovationsgrad oder Marktpotential abgefragt wurden und das damit in weiten Teilen der Expertenbewertung entspricht (vgl. Kapitel 3.3 und Anhang). Die beiden Skalen sind in den Abbildungen 1 und 2 dargestellt. Es erfolgte keine Zuordnung der Teilnehmer auf die zu bewertenden Ideen. Jede Idee konnte durch beide Skalen und durch jeden Teilnehmer bewertet werden. Dabei wurde jedoch sichergestellt, dass jede Idee von jedem Teilnehmer nur einmal mit

einer der beiden Bewertungsskalen bewertet werden konnte. Die Bewertung der Ideen war während der gesamten Wettbewerbsdauer möglich. Den Teilnehmern erhielten über die Bewertungsskalen hinaus keine weitere Informationen. Durchschnittlich wurde jede Idee mit der binären Skala 3,7 mal und der komplexen Skala 2,6 mal bewertet.



Abbildung 1: binäre Skala zur Messung von Ideenqualität

| | | | | | |
|---|----------------------|-----------|---------|-----------------|--------------------|
| Originalität "Die Idee ist neuartig und eigenllich." | stimme voll zu | stimme zu | neutral | stimme nicht zu | stimme garnicht zu |
| Kommentar: | <input type="text"/> | | | | |
| Innovationsgrad "Die Idee ist eine Neukombination von Faktoren, die sich wirtschaftlich verwerten lässt." | stimme voll zu | stimme zu | neutral | stimme nicht zu | stimme garnicht zu |
| Kommentar: | <input type="text"/> | | | | |
| Kundennutzen "Die Idee ist zweckgemäß und schafft einen Mehrwert, der sich wirtschaftlich kommunizieren lässt." | stimme voll zu | stimme zu | neutral | stimme nicht zu | stimme garnicht zu |
| Kommentar: | <input type="text"/> | | | | |
| Nutzerakzeptanz "Ein bestehender Bedarf wird durch die Idee gedeckt." | stimme voll zu | stimme zu | neutral | stimme nicht zu | stimme garnicht zu |
| Kommentar: | <input type="text"/> | | | | |
| Realisierbarkeit "Die Umsetzung der Idee ist wirtschaftlich möglich." | stimme voll zu | stimme zu | neutral | stimme nicht zu | stimme garnicht zu |
| Kommentar: | <input type="text"/> | | | | |
| Kundenpotential (Marktgröße) "Die zu erwartende Nachfrage der Zielgruppe rechtfertigt die Umsetzung der Idee." | stimme voll zu | stimme zu | neutral | stimme nicht zu | stimme garnicht zu |
| Kommentar: | <input type="text"/> | | | | |
| Vermarktungsfähigkeit "Die Idee passt zu SAP und ist konkurrenzfähig" | stimme voll zu | stimme zu | neutral | stimme nicht zu | stimme garnicht zu |
| Kommentar: | <input type="text"/> | | | | |

Abbildung 2: Mehrdimensionale Skala zur Messung von Ideenqualität

3.3 Evaluationsdesign

Zur Untersuchung der Validität der eingesetzten Bewertungsskalen wurde die wahre, jedoch unbekannte Ideenqualität mittels einer umfangreichen Expertenbewertung angenähert. Dabei wurde auf die Consensual Assessment Technique (Amab96) zurückgegriffen, die bereits erfolgreich für die Bewertung von Kundengenerierten Produktideen verwendet wurde (KrGA04; FrHi06; Walc07). Bei dieser Methode werden die Ideen durch eine Gruppe von Experten unabhängig von einander, auf einer Rating-Skala bewertet (im vorliegenden Fall 1 (sehr schlecht) bis 7 (sehr gut)). Dafür wurde ein aus 15 Items bestehender Evaluierungsbogen entwickelt (vgl. Anhang). Die Jury bestand aus 6 Mitarbeitern von SAP und den SAP University Competence Centern München und Magdeburg. Zur Reliabilitäts- und Validitätssicherung wurden die Expertenbewertungen anschließend einer Faktorenanalyse unterzogen, in deren Anschluss aus den verbleibenden Items ein additiver Qualitätsindex gebildet wurde.

Die Teilnehmerbewertung ist von hoher Qualität, wenn die Teilnehmer in der Lage sind, unter den eingereichten Ideen die besten effektiv herauszufiltern. Da diese Bewertung letztendlich aber immer nur eine Vorauswahl für eine weitere interne Begutachtung beim Veranstalter darstellen kann, ist für diesen das genaue Qualitätsmaß im Prinzip irrelevant. Es zählt nur, dass die besten Ideen zielsicher identifiziert werden. Eine mäßige oder bereits bekannte Idee ist für den Veranstalter gleichermaßen uninteressant (ReBN07).

Die Validität einer Messung bezeichnet allgemein die Gültigkeit einer Messung, d.h. sie gibt an, in welchem Grad ein Messinstrument misst, was es zu messen vorgibt. Zur Bestimmung der Validität der beiden Skalen zur Teilnehmerbewertung wurde in einem ersten Schritt jeweils ein Qualitätsindex für die beiden Skalen gebildet. Bei der binären Skala wurde hierfür der Differenzwert (positive Bewertungen („Go“) abzüglich negativer Bewertungen („NoGo“)) verwendet, bei der mehrdimensionalen Skala die einzelnen Items gemittelt. In einem zweiten Schritt wurden die Qualitätsindexe der Skalen zur Teilnehmerbewertung sowie der Expertenbewertung dichotomisiert. Verschiedene Arbeiten über von Kunden

generierten Produktideen zeigen, dass unter diesen ca. 10-30% als qualitativ hochwertig einzuschätzen sind (FrHi06; Walc07). Aus diesem Grund wurden schrittweise jeweils die besten 10%, 20% und 30% der eingereichten Ideen als „gut“ und die restlichen als „schlecht“ klassifiziert. Mittels Kreuztabellen wurde anschließend die Skalen mit der ebenfalls dichotomisierten Expertenbewertung verglichen und so deren Bewertungsgüte untersucht. Im Wesentlichen handelte es sich hierbei um die Bestimmung der Übereinstimmungsvalidität zwischen den beiden Bewertungsskalen und der Expertenbewertung.

4 Empirische Ergebnisse

4.1 Validierung der Expertenbewertung

Als Vergleichsmaßstab für die Bewertungsgüte der getesteten Bewertungsskalen diente die unabhängige Expertenbewertung, die als valide angenommen werden kann, wenn für das verwendete Messinstrument Konstruktvalidität angenommen werden kann. Konstruktvalidität kann mittels Faktoranalysen überprüft werden und liegt vor, wenn für die zu testenden Faktoren konvergente (d.h. alle zu einem Faktor gehörigen Items laden nur auf diesen) und diskriminante Validität (d.h. die einzelnen Faktoren messen unterschiedliche Aspekte des zu validierenden Konstrukts) vorliegt (Bühn08).

Zur Validitätssicherung der Expertenbewertung wurden die Items in einem ersten Schritt einer exploratorischen Faktorenanalyse mit der Statistiksoftware SPSS 16 unterzogen. Die MSA-Maße für alle Items und die gesamte Datenstruktur waren $>0,6$ (Bühn08). Jedoch zeigten die Items N6, R1, C3 und F3 hohe Querladungen, so dass eine Elimination dieser Items notwendig war. Mit einem Cronbach'schen Alpha von min. 0.841 wiesen alle Faktoren die geforderte Mindestreliabilität von 0,7 auf (Bühn08). Anschließend wurden die Faktoren einer konfirmatorischen Faktorenanalyse mit der Statistiksoftware Amos 4.01 unterzogen. Die Faktoren Neuartigkeit, Relevanz und Konkretisierung zeigten sehr hohe Faktorreliabilitäten und sehr hohe Werte für die durchschnittliche erfasste Varianz (DeV) (vgl. Tabelle 1). Faktorreliabilitäten von $>0,6$ und Durchschnittlich erfasste Varianzen von

>0,5 können hier als Grenzwerte für eine qualitativ hochwertige Messung angesehen werden (HoBa95). Folglich kann für diese Faktoren Konvergenzvalidität angenommen werden. Der Faktor Umsetzbarkeit erfüllte diese jedoch nicht und musste daher aus der weiteren Analyse ausgeschlossen werden.

| Faktor | Item | Faktorladungen | | Eigenwerte | α | Indikatorreliabilität | Faktorreliabilität | DeV |
|-----------------|------|----------------|--------------|--------------|----------|-----------------------|--------------------|------|
| Neuartigkeit | N1 | 0,965 | 0,167 | -0,049 | 5,575 | 0,956 | 0,853 | 0,83 |
| | N2 | 0,890 | 0,255 | -0,101 | | | 0,825 | |
| | N3 | 0,853 | 0,257 | -0,087 | | | 0,938 | |
| | PR1 | 0,839 | 0,220 | -0,065 | | | 0,879 | |
| | PR2 | 0,706 | 0,210 | -0,283 | | | 0,604 | |
| Relevanz | R2 | 0,281 | 0,909 | -0,084 | 2,373 | 0,841 | 0,879 | 0,95 |
| | R3 | 0,359 | 0,846 | -0,150 | | | 0,794 | |
| Konkretisierung | A1 | -0,235 | -0,194 | 0,912 | 1,118 | 0,887 | 0,896 | 0,91 |
| | A2 | -0,064 | -0,028 | 0,760 | | | 0,762 | |

Tabelle 1: Ergebnisse der Faktorenanalyse

Die Diskriminanzvalidität wurde auf Basis des Fornell-Larcker-Kriteriums bestimmt, das fordert, dass die DeV eines Faktors größer ist als die quadrierte Korrelation mit jedem anderem Faktor. Tabelle 2 zeigt, dass für jeden der verbleibenden Faktoren Diskriminanzvalidität angenommen werden kann. Alle Items übertrafen eine Mindestindikatorreliabilität von 0,4 (HoBa95) – die gute Reliabilität der Faktoren konnte hiermit bestätigt werden.

| | Neuartigkeit | | Relevanz |
|------------------------|--------------|--------------------------|----------|
| | DeV | Quadrierte Korrelationen | |
| Neuartigkeit | 0,83 | | |
| Relevanz | 0,90 | 0,04 | |
| Konkretisierung | 0,83 | 0,08 | 0,02 |

Tabelle 2: Diskriminanzvalidität der Faktoren

Abschließend wurde der globale Fit getestet. Der χ^2 -Test war signifikant, aber das χ^2/df -Verhältnis lag mit 2,425 unter dem geforderten Maximalwert von 2,5 (HoBa95). Auch die anderen Fit Indizes zeigten einen guten Fit (Sollwert in

Klammern): GFI=0,82 (>0,9), CFI=0,94 (>0,9), NFI=0.90 (>0,9) und SRMR=0,04 (<0,11) (Bühn08).

4.2 Bewertungsgüte der Teilnehmerbewertung

Betrachtet man die Kreuztabellen in Tabelle 3 und die Ergebnisse der Kontingenzanalyse in Tabelle 4, wird sofort ersichtlich, dass sich bei den Cut-Off-Niveaus von 20% und 30% für die binäre Skala ein Zusammenhang zwischen Teilnehmer- und Expertenbewertung ergibt. Beim Vergleich der beiden Skalen wurden jeweils mehr Ideen richtig als „gut“ bzw. „schlecht“ klassifiziert, als bei statistischer Unabhängigkeit zu erwarten wäre. In diesen beiden Fällen stimmt die Teilnehmerbewertung gut mit der Expertenbewertung überein. Die χ^2 -Tests sind mit $p < 0,01$ hoch signifikant. Auch die weiteren Assoziationsmaße sind signifikant und liegen deutlich über dem geforderten Mindestniveau von 0,3 (BEPW06). Der stärkste Zusammenhang ergibt sich bei der binären Skala bei einem Niveau von 20%. Hier wurden von den 11 „guten“ Ideen der Expertenbewertung 6 von den Teilnehmern richtig als solche erkannt. Das sind 4 mehr als sich bei statistischer Unabhängigkeit bzw. zufälligem Auswählen der Ideen ergeben hätten. Ein anderes Bild ergibt sich jedoch für die mehrdimensionale Skala: Hier ergibt sich bei keinem der drei Cut-Off-Niveaus ein signifikanter Zusammenhang.

| | | Binäre Skala | | | Mehrdimensionale Skala | | | |
|------------|----------|------------------|------|----------|------------------------|------|----------|----|
| | | Schlecht | Gut | Σ | Schlecht | Gut | Σ | |
| Top 10% | Schlecht | Anzahl | 48 | 4 | 52 | 46 | 6 | 52 |
| | | Erwartete Anzahl | 47,4 | 4,6 | 52 | 46,5 | 5,5 | 52 |
| | Gut | Anzahl | 4 | 1 | 5 | 5 | 0 | 5 |
| | | Erwartete Anzahl | 4,6 | 0,4 | 5 | 4,5 | 0,5 | 5 |
| Top 20% | Schlecht | Anzahl | 41 | 5 | 46 | 37 | 10 | 46 |
| | | Erwartete Anzahl | 37,1 | 8,9 | 46 | 37,9 | 9,1 | 46 |
| | Gut | Anzahl | 5 | 6 | 11 | 9 | 1 | 11 |
| | | Erwartete Anzahl | 8,9 | 2,1 | 11 | 8,1 | 1,9 | 11 |
| Top 30% | Schlecht | Anzahl | 30 | 10 | 40 | 31 | 9 | 40 |
| | | Erwartete Anzahl | 25,3 | 14,7 | 40 | 28,1 | 11,9 | 40 |
| | Gut | Anzahl | 6 | 11 | 17 | 9 | 8 | 17 |
| | | Erwartete Anzahl | 10,7 | 6,3 | 17 | 11,9 | 5,1 | 17 |

Tabelle 3: Kreuztabellen

| | | χ^2 (Pearson) | Phi | Cramers V | Kontingenzkoeffizient |
|------------|------------------------|--------------------|---------|-----------|-----------------------|
| Top | Binäre Skala | 0,863 | 0,123 | 0,123 | 0,123 |
| 10% | Mehrdimensionale Skala | 0,645 | 0,106 | 0,106 | 1,06 |
| Top | Binäre Skala | 10,873** | 0,437** | 0,437** | 0,400** |
| 20% | Mehrdimensionale Skala | 0,673 | 0,109 | 0,109 | 0,108 |
| Top | Binäre Skala | 8,083** | 0,377** | 0,377** | 0,352** |
| 30% | Mehrdimensionale Skala | 3,438 | 0,246 | 0,246 | 0,239 |

** signifikant mit $p < 0,01$

Tabelle 4: Ergebnisse der Kontingenzanalyse

Aus diesen Ergebnissen wird ersichtlich, dass im vorliegenden Ideenwettbewerb die einfache binäre Skala dem komplexeren Bewertungsformular überlegen war und eine höhere Übereinstimmungsvalidität mit der Expertenbewertung aufwies.

5 Diskussion und Ausblick

Im Laufe der Zeit wurden eine Vielzahl von mehrdimensionalen Bewertungsskalen zur Bewertung von Ideen und Innovationen entwickelt, die valide und reliable Qualitätseinschätzungen erlauben. Die meisten dieser Bewertungsinstrumente wurden jedoch für Bewerter mit einer hohen Expertise bzgl. der Bewertungsobjekte entwickelt. Daher kann prinzipiell in Frage gestellt werden, ob diese Bewertungsinstrumente auch für Teilnehmerbewertungen in Ideenwettbewerben geeignet sind. Bewertungen sind immer subjektiv und die Wahl der richtigen Bewertungsskala hängt im erheblichen Maße von der Differenzierungsfähigkeit der Probanden ab, die von deren Intelligenz, Bildung und Vertrautheit mit dem zu untersuchenden Sachverhalt bestimmt wird (Stie98). Daher kann grundsätzlich davon ausgegangen werden, dass (Unternehmens)-Experten in der Lage sind feiner zu differenzieren als ein allgemeines Publikum wie z.B. die Teilnehmer des Ideenwettbewerbs. Weiterhin bedarf eine adäquate Beurteilung vieler Bewertungspunkte des komplexen Bewertungsformulars, wie z.B. die Umsetzbarkeit oder die Vermarktbarkeit einer Idee, einer Kenntnis von unternehmensinterner Abhängigkeiten des Veranstalters. Jedoch sind Punkte, wie z.B. die Unterneh-

mensstrategie, die Geschäftsprozesse oder die verfügbaren Ressourcen, den Teilnehmern im Detail unbekannt.

Im vorliegenden Fall könnte die Fehleinschätzung der Ideenqualität mittels der komplexeren Skala durch eine Überforderung der Teilnehmer hervorgerufen worden sein. Es ist zu vermuten, dass durch ein zu detailliertes Bewertungsformular einzelne Qualitätsaspekte der Ideen falsch bewertet wurden, obwohl die Teilnehmer grundsätzlich in der Lage gewesen wären die Qualität richtig einzuschätzen. Weiterhin kann angenommen, dass die Motivation der Bewerter und die Gründlichkeit, mit der die Bewertung durchgeführt werden, ebenfalls wesentliche Einflussfaktoren für die Qualität der Ideenbewertung darstellen. Komplexe Bewertungsformulare bedürfen aber einer höheren kognitiven Verarbeitungstiefe und bergen daher die Gefahr nicht „richtig“ ausgefüllt zu werden. Zur Nutzbarmachung der kollektiven Intelligenz der Ideengeber scheint daher das Abfragen des bloßen „Bauchgefühls“ die bessere Wahl zu sein. Eine Möglichkeit zur Verbesserung der User Ratings könnte jedoch in der Verankerung der Skalen liegen. Eine Unterstützung der Teilnehmer in Form von Skalenbeschriftungen, wie z.B. „sofort realisierbar“ oder „realisierbar in zwei Jahren“, könnte im Gegensatz zu einer abstrakten Bewertung der Realisierbarkeit von 1 bis 5 zu einer Verbesserung der Bewertungsgüte führen.

Weiterer Forschungsbedarf ergibt sich jedoch nicht nur in der Wahl der richtigen Skalierung. Es ist weitgehend unbekannt, welche Einflussfaktoren die Unterschiede zwischen Experten- und Teilnehmerbewertung hervorrufen. Neben den Fähigkeiten und Eigenschaften der Bewerter ist hier vor allem der Einfluss eines möglichen Bewertungsbias interessant, der sich durch opportunistisches Teilnehmerverhalten oder ungünstiges Plattform-Design ergeben könnte. Einerseits könnten Teilnehmer fremde Ideen absichtlich schlecht bewerten, um die eigenen Siegchancen zu erhöhen. Deshalb wäre an dieser Stelle bspw. zu untersuchen, ob sich durch Abschwächung der Konkurrenzsituation eine Verbesserung der Teilnehmerbewertung erreichen ließe. Andererseits werden in Ideenwettbewerben und Innovationscommunities gut bewertete Ideen in der Regel prominent auf der entsprechenden

Interaktionsplattform platziert, oftmals direkt auf der Startseite, wodurch sie mehr Bewertungen erhalten könnten und die Bewertung verzerrt würde.

Literatur

- [Amab96] Amabile, T. M., 1996, Creativity in context. Update to social psychology of creativity, Oxford.
- [AvZe97] Avery, C., Zeckhauser, R., 1997, Recommender Systems for Evaluating Computer Messages, in: Communications of the ACM, Vol: 40, Nr. 3, S. 88-89.
- [BEPW06] Backhaus, K., Erichson, B., Plinke, W., Weiber, R., 2006, Multivariate Analysemethoden: Eine Anwendungsorientierte Einführung, Berlin.
- [BrLK07] Bretschneider, U., Ebner, W., Leimeister, J. M., Krcmar, H., 2007, Internetbasierte Ideenwettbewerbe als Instrument der Integration von Kunden in das Innovationsmanagement von Software-Unternehmen, Proceedings der Gemeinschaften in Neuen Medien 2007 (GeNeMe'07), Dresden, 1.-2.10.2007.
- [BrLK09] Bretschneider, U., Leimeister, J. M., Krcmar, H., (2009), Methoden der Kundenintegration in den Innovationsprozess: Eine Bestandsaufnahme, Arbeitspapier des Lehrstuhl für Wirtschaftsinformatik, Technische Universität München, Vol.: 34.
- [Bühn08] Bühner, M., 2008, Einführung in die Test- und Fragebogenkonstruktion, München et. al.
- [CaVa99] Cady, S. H., Valentine, J., 1999, Team Innovation and perceptions of consideration. What difference does diversity make?, in: Small Group Research, Vol: 30, Nr. 6, S. 730-750.
- [DHRS06] Dean, D. L., Hender, J. M., Rodgers, T. L., Santanen, E. L., 2006, Identifying quality, novel, and creative Ideas: Constructs and scales for idea evaluation, in: Journal of the Association for Information Systems, Vol: 7, Nr. 10, S. 646-698.
- [ELBK08] Ebner, W., Leimeister, M., Bretschneider, U., Krcmar, H., 2008, Leveraging the wisdom of crowds: Designing an IT-supported ideas competition for an ERP software company, Proceedings der 41st Annual Hawaii International Conference on System Sciences (HICCS 41), Big Island, Hawaii, 7.-10.01.2008.
- [FiWS96] Finke, R. A., Ward, T. B., Smith, S. M., 1996, Creative cognition. Theory, research and applications, Cambridge.
- [FrHi06] Franke, N., Hienerth, C., 2006, Prädikatoren der Qualität von Geschäftsideen: Eine empirische Analyse eines Online-Ideen-Forums, in: Zeitschrift für Betriebswirtschaft Special Issue Vol: 6, Nr. 4, S. 47-68.

-
- [HoBa95] Homburg, C., Baumgartner, H., 1995, Beurteilung von Kausalmodellen. Bestandsaufnahme und Anwendungsempfehlungen, in: Marketing Zeitschrift für Forschung und Praxis, Vol: 17, Nr. 3, S. 162-176.
- [Kais09] Kaiserswerth, 2009, InnovationJam: Unterstützung globaler Zusammenarbeit und Innovation, in: A. Picot and S. Doebelin, Innovationsführerschaft durch Open Innovation: Chancen für die Telekommunikation-, IT- und Medienindustrie, Berlin, 155-170.
- [KrGA04] Kristensson, P., Gustafsson, A., Archer, T., 2004, Harnessing the creative potential among users, in: The Journal of Product Innovation Management, Vol: 21, Nr. 1, S. 4-14.
- [LHBK09] Leimeister, J. M., Huber, M., Bretschneider, U., Krcmar, H., 2009, Leveraging Crowdsourcing - Activation-Supporting Components for IT-based Idea Competitions, in: Journal of Management Information Systems, Vol: 26, Nr. 1, in press.
- [MaWa94] MacCrimmon, K. R., Wagner, C., 1994, Stimulating ideas through creative software, in: Management Science, Vol: 40, Nr. 11, S. 1514-1532.
- [NaBo94] Nagasundaram, M., Bostrom, R. P., 1994, The structuring of creative processes using GSS: A framework for research, in: Journal of Management Information Systems, Vol: 11, Nr. 3, S. 87-114.
- [Nich97] Nichols, D. M., 1997, Implicit Rating and Filtering, Proceedings der 5th DELOS Workshop on Filtering and Collaborative Filtering, Budapest Hungary, 10-12.11.1997.
- [PIBD04] Plucker, J. A., Beghetto, R. A., Dow, G. T., 2004, Why isn't creativity more important to educational psychologists? Potentials, pitfalls, and future directions in creativity research, in: Educational Psychologist, Vol: 39, Nr. 2, S. 83-96.
- [RePi06] Reichwald, R., Piller, F., 2006, Interaktive Wertschöpfung, Wiesbaden.
- [ReBN07] Reinig, B. A., Briggs, R. O., Nunamaker Jr, J. F., 2007, On the measurement of ideation quality, in: Journal of Management Information Systems, Vol: 23, Nr. 4, S. 143-161.
- [Roch91] Rochford, L., 1991, Generating and screening new product ideas, in: Industrial Marketing Management, Vol: 20, Nr. 4, S. 287-296.
- [Soll06] Soll, J. H., 2006, Ideengenerierung mit Konsumenten im Internet, Wiesbaden.
- [Stie98] Stier, W., 1998, Empirische Forschungsmethoden, Berlin.
- [Walc07] Walcher, P.-D., 2007, Der Ideenwettbewerb als Methode der aktiven Kundenintegration, Wiesbaden.
- [WHPT09] Winkelmann, A., Herwig, S., Pöppelbuß, J., Tiebe, D., Becker, J., 2009, Discussion of Functional Design Options for Online Rating Systems: A State-of-the-Art Analysis, Proceedings der 17th European

Anhang

- N1 Die Idee ist neuartig?
- N2 Die Idee ist rar bzw. einzigartig?
- N3 Die Idee ist ungewöhnlich, fantasievoll oder überraschend?
- PR1 Die Idee ist revolutionär?
- PR2 Die Idee ist radikal?
- PR3 Die Idee ist trendsetzend?
- R1 Die Idee erfüllt einen klar definierten Kundennutzen?
- R2 Die Idee ermöglicht für SAP das Erschließen eines attraktiven Marktpotenzials?
- R3 Die Idee ermöglicht für SAP das Schaffen von langfristigen, schwer imitierbaren strategischen Wettbewerbsvorteilen?
- U1 Die Idee ist technisch leicht realisierbar?
- U2 Die Idee lässt sich wirtschaftlich umsetzen?
- U3 Die Idee passt zum Markenimage von SAP?
- A1 Die Idee ist exakt, vollständig und präzise beschrieben?
- A2 Die Idee ist ausgereift?
- A3 Die Idee bzw. ihr Nutzen wird klar und überzeugend kommuniziert?