

# Konstruktion eines Fragebogens zur Messung der User Experience von Softwareprodukten

Bettina Laugwitz, Martin Schrepp, Theo Held

SAP AG

## **Zusammenfassung**

Mit dem User Experience Questionnaire wurde ein Fragebogen entwickelt, der eine schnelle Messung verschiedener Kriterien der Softwarequalität erlaubt. Die Relevanz der Kriterien für die Beurteilung wurde durch eine empirische Selektion sichergestellt. Experten sammelten und reduzierten eine große Menge potenziell relevanter Begriffe und Aussagen, die sowohl „harte“ Usability-Kriterien als auch „weichere“ User Experience-Kriterien einschlossen. Der daraus entstandene ursprüngliche Fragebogen mit bipolaren 80 Items wurde in mehreren Untersuchungen eingesetzt und durch eine Faktorenanalyse auf 26 Items reduziert, die sich den sechs Faktoren Attraktivität, Durchschaubarkeit, Effizienz, Vorhersagbarkeit, Stimulation und Originalität zuordnen lassen. Erste Validierungsuntersuchungen deuten auf eine zufriedenstellende Konstruktvalidität hin.

## 1 Einleitung

Benutzerfragebögen zur Usability sind unter Umständen als alleinige Methode zur Evaluation ausgelegt und geeignet, wie der IsoMetrics (Gediga & Hamborg 1999; auch Hamborg 2002). Zumeist sind sie aber nur im Zusammenspiel mit weiteren Erhebungsmethoden sinnvoll einzusetzen (vgl. Dzida et al. 2000). Sie können dabei als eher grober Indikator für Produkteigenschaften dienen (z.B. AttrakDiff2, Hassenzahl et al. 2003) oder auch Hinweise auf konkrete Gebrauchstauglichkeitsprobleme liefern (z.B. SUMI, s. Kirakowski & Corbett 1993). Die Ergebnisse müssen aber immer im Zusammenhang mit anderen Ergebnissen betrachtet oder durch Experten interpretiert werden (vgl. Dzida et al. 2000).

Zusätzlich zu Methoden, die differenzierte Beurteilungen von Produktstärken und –schwächen liefern, wie ein Benutzertest oder eine Expertenbeurteilung mittels Heuristischer Evaluation (Nielsen 1994), kann ein quantitatives Maß der Benutzerbeurteilung hilfreich sein, bevorzugt in Form eines Fragebogens. Man kann dem Benutzer das Feedback erleichtern, indem man ihm erlaubt, genau das zu äußern, was ihm bei der Beurteilung besonders

nahe liegt: Wie hat er die Software und seine Interaktion damit erlebt? Damit müssen nicht nur Aspekte gemeint sein, wie sie durch die ISO 9241-10 oder durch die Kriterien der Effektivität und Effizienz gemäß ISO 9241-11 beschrieben sind. Gerade die etwas diffuseren Qualitätskriterien, die den User Experience Goals nach Preece et al. (2002) entsprechen und sich beispielsweise in der hedonischen Qualität des AttrakDiff2 (Hassenzahl et al. 2003) oder dem Kriterium der Benutzerzufriedenheit nach ISO 9241-11 widerspiegeln, sind geeigneter Gegenstand eines Benutzerfragebogens.

Ziel der unten beschriebenen Fragebogenentwicklung war es, die schnelle Erhebung eines umfassenden Gesamteindrucks der User Experience (z.B. Preece et al. 2002) aus Sicht des Benutzers zu ermöglichen, der unmittelbar und einfach das beschreibt, was der Benutzer beim Umgang mit dem Produkt empfunden und wie er das Produkt erlebt hat. Diese Zielsetzung gründet sich auf folgende Vorüberlegungen:

**Schnelle Erhebung:** Fragebögen sind besonders ökonomisch in Anwendung und Auswertung. Manche Fragebögen sind dennoch in der absoluten Durchführungszeit relativ aufwändig. Ein Beispiel ist der SUMI (Kirakowski & Corbett 1993), bei dem der Benutzer seine Zustimmung zu 50 Aussagen zur Gebrauchstauglichkeit äußern soll, oder die Langform des IsoMetrics (Gediga & Hamborg 1999), die ein Rating von 75 Items vorsieht. Diese Fragebögen sollen eine komplette Beurteilung der Gebrauchstauglichkeit einschließlich differenzierter Problembeschreibungen ermöglichen und als alleinige Gebrauchstauglichkeitsmaße verwendet werden können (s. z.B. Hamborg 2002). Dieser Anspruch soll vom neuen Fragebogen nicht erfüllt werden, da er als ergänzende Methode zusätzlich zu Heuristischen Evaluationen oder Benutzertests eingesetzt werden soll.

**Umfassender Gesamteindruck zum Produkterleben:** Traditionelle Methoden legen ihren Schwerpunkt meist auf Usability-Kriterien im engeren Sinne. Diese entsprechen etwa den Usability Goals, wie Preece et al. (2002) sie beschreiben, oder auch der pragmatischen Qualität gemäß Hassenzahl et al. (2003). Neuere Ansätze fokussieren immer mehr auf das subjektive, auch emotionale Erleben des Benutzers, was in der ISO 9241-11 unter dem Kriterium der Benutzerzufriedenheit eingeordnet werden kann. Preece et al. (2002) bezeichnen diese Kriterien als User Experience Goals. Hassenzahl et al. (2003) berücksichtigen sie unter dem Aspekt der hedonischen Qualität.

Zur Beurteilung der pragmatischen und hedonischen Eigenschaften auch von betriebswirtschaftlicher Software durch den Benutzer wurde der AttrakDiff2 bereits mit Erfolg eingesetzt (s. Schrepp et al. 2004). Allerdings liegt beim AttrakDiff2 der Schwerpunkt auf den hedonischen Aspekten der Qualität, was für eine umfassende Bewertung der Softwarequalität z.B. von professioneller betriebswirtschaftlicher Software nicht unbedingt optimal ist. Der SUMI (Kirakowski & Corbett 1993) kann als Beispiel für ein anderes Extrem gelten: nur eine von fünf Skalen zielt auf die Erfassung emotionaler Aspekte.

Ein umfassender Gesamteindruck muss sicherlich alle Aspekte und Produkteigenschaften einschließen, die für den Benutzer von Relevanz sind. Für den Fragebogen sollen daher a priori weder pragmatische noch hedonische Kriterien ausgeschlossen oder bevorzugt werden. Die potenziellen Items sollten möglichst breitgefächert sein und erst durch empirische Daten mithilfe einer explorativen Faktorenanalyse ausgewählt und gruppiert werden.

**Unmittelbar und einfach:** Wie fühlt sich die Interaktion mit dem Produkt an? Welche Produkt- und Interaktionseigenschaften sind dem Benutzer besonders aufgefallen? Der Benutzer

soll möglichst unmittelbar, spontan und ohne tiefgehende rationale Analyse seine Beurteilung über das Produkt äußern können. Dieser Ansatz wird beispielsweise auch vom AttrakDiff2 verfolgt (s. Hassenzahl et al. 2003).

Das durch den neuen Fragebogen zu erhebende Benutzerfeedback soll im Normalfall nicht die einzige Informationsquelle der Beurteilung der Softwarequalität darstellen, sondern z.B. einen Benutzertest ergänzen. Der Benutzer soll nicht von seinem Erleben der Interaktion abstrahieren oder sich an womöglich vergessene oder übersehene Details erinnern müssen. Die explizite und nachträgliche Beurteilung durch den Benutzer ist nicht immer verlässlich, wie z.B. Nielsen (2001) anmerkt. So zeigen Befunde von Laugwitz (2001), dass sich farblich unterscheidende Benutzeroberflächen zwar unterschiedliche Wirkungen auf das Erleben der Benutzer haben (z.B. auf ihre Stimmung). Dieser Unterschied spiegelt sich aber nicht in den Antworten auf Fragen zur Einschätzung des UIs wider.

Experten können Benutzeroberflächen detailliert bewerten, Benutzer können bei der Interaktion mit dem Produkt beobachtet werden. Daher kann der Fokus eines Fragebogens auf Kriterien liegen, die dem Benutzer unmittelbar zugänglich sind, nämlich auf der subjektiven Wahrnehmung von Produkteigenschaften und von deren Einfluss auf den Benutzer selbst. Dieser Anspruch ähnelt dem von Hassenzahl et al. (2003), weshalb die Erstellung eines semantischen Differentials vom Format des AttrakDiff2 angestrebt wurde und die methodische Vorgehensweise von Hassenzahl et al. als Vorbild diente.

## 2 Konstruktion des Fragebogens

### 2.1 Erzeugung des Itempools

In zwei Brainstorming Sitzungen (Dauer jeweils ca. 1,5 Stunden) wurden insgesamt 15 bei der SAP AG angestellte Usability Experten gebeten, Vorschläge für deutschsprachige Begriffe und Aussagen zu machen, die nach ihrer Ansicht charakteristisch für die Einschätzung von Benutzererleben (User Experience) seien. Die Sitzungen wurden moderiert, die Vorschläge wurden von der Moderatorin bzw. vom Moderator protokolliert.

Den Experten wurden die folgenden Fragen gestellt: (1) „Auf welche Produkteigenschaften reagieren Nutzer besonders intensiv?“, (2) „Welche Gefühle oder Einstellungen rufen Produkte bei Nutzern hervor?“ und (3) „Wie sind die typischen Reaktionen von Nutzern während oder nach einer Untersuchung zur Gebrauchstauglichkeit?“ Die gesammelten Vorschläge (insgesamt 229) wurden anschließend konsolidiert, redundante Begriffe wurden entfernt. Sofern ein Listeneintrag noch nicht als Adjektiv vorlag, wurde er durch das zum jeweiligen Begriff passende Adjektiv ersetzt. Die bereinigte Liste enthielt 221 Adjektive. Sieben Usability Experten (allesamt Teilnehmer in einer der Brainstorming Sitzungen) selektierten anschließend jeweils 25 bevorzugte Adjektive („Top 25“) und vergaben Vetos für Adjektive, die sie für absolut ungeeignet hielten (ohne Begrenzung der Anzahl). Adjektive, die mehr als ein Veto erhielten oder weniger als zweimal in einer der Top 25 Listen auftauchten, wurden aus der Gesamtliste entfernt. Übrig blieben 80 Adjektive, die den oben genannten Kriterien genügten. Da für den Fragebogen das Format eines semantischen Differentials geplant war,

wurde für jedes Adjektiv das bestpassende Antonym bestimmt. Die Liste der Adjektivpaare wurde in eine zufällige Reihenfolge gebracht. Außerdem wurde eine zweite Version der Liste mit umgekehrter Reihenfolge und jeweils komplementären Polungen der Paare erstellt. Die Darstellung erfolgte in Form eines siebenstufigen semantischen Differentials:

Attraktiv ①      ②      ③      ④      ⑤      ⑥      ⑦ unattraktiv

## 2.2 Datenerhebung zur Vorbereitung der Itemreduktion

Um die spezifischen Eigenschaften der Adjektivpaare in Hinblick auf die Beurteilung von Softwareprodukten zu untersuchen, wurden die beiden Versionen des (Roh-)Fragebogens in sechs Untersuchungen vorgegeben (siehe Tabelle 1).

Tabelle 1: Untersuchungen zur Vorbereitung der Itemreduktion

	Bezeichnung	Ort	Typ	Anzahl Teilnehmer	
				Version 1	Version 2
1	SYSTAT	UNI Mannheim	Paper/Pencil	13	14
2	HANDY	UNI Mannheim	Paper/Pencil	23	25
3	BSCW	UNI Mannheim	Paper/Pencil	7	7
4	SELECTION	UNI Mannheim	Paper/Pencil	13	13
5	MOBILE	SAP Walldorf	Paper/Pencil	8	7
6	PCC	SAP Walldorf	Online	12	11
				$\Sigma$ 76	$\Sigma$ 77

Erläuterungen zu den einzelnen Untersuchungen:

1. SYSTAT: Die Teilnehmer eines Einführungskurses in die Statistik-Software SYSTAT wurden gebeten, eine vorgegebene Aufgabenstellung mit SYSTAT zu bearbeiten, bzw. den Aufgabenbearbeiter zu beobachten. Anschließend wurden die Teilnehmer gebeten, einen der beiden Fragebogen in Hinblick auf die gerade durchlaufene Aufgabenstellung auszufüllen.
2. HANDY: Die Teilnehmer eines Psychologie-Seminars wurden gebeten, einen Eintrag in das Adressbuch ihres Mobiltelefons zu machen und diesen Eintrag anschließend wieder zu löschen. Mit Hilfe des Fragebogens sollten sie dann die Bedienbarkeit des Mobiltelefons in Bezug auf die gerade durchlaufenen Bedienschritte beurteilen.
3. BSCW: Die Hörer einer Vorlesung sollten das innerhalb der Lehrveranstaltung verwendete Programm zur Online-Kollaboration BSCW beurteilen. Jeder der Teilnehmer hatte vor dem Ausfüllen des Fragebogens aktiv mit dem beurteilten Programm gearbeitet.
4. SELECTION: Die Teilnehmer eines Informatik-Seminars wurden gebeten, wahlweise eines der folgenden Software-Produkte zu beurteilen: Eclipse Development Workbench,

Borland JBuilder, Microsoft Visual Studio, Mozilla 1.7 Browser, Microsoft Internet Explorer 6, sowie den Firefox Browser in der Version 1.0. Bewertungen wurden abgegeben für Firefox 1.0, Microsoft Internet Explorer 6 und die Eclipse Workbench.

5. MOBILE: Im Rahmen eines regelmäßigen Treffens von Usability Experten der SAP AG wurde eine Variante der Benutzungsschnittstelle der SAP Customer Relationship (CRM) Software vorgestellt. Die anwesenden Experten wurden nach der Präsentation gebeten, einen der beiden Fragebögen auszufüllen.
6. PCC: Mit Hilfe des bei der SAP gebräuchlichen „Enduser Feedback Service“ wurde eine Online-Befragung zu einer weiteren Variante der SAP CRM Software vorbereitet. Die Befragung bestand aus einer kurzen Präsentation eines typischen Interaktionsablaufes mit der Software und einer anschließenden Darbietung des Fragebogens. Die Zugangsdaten für die Befragung wurden an alle SAP Usability Experten am Standort Walldorf versandt. Die Zuordnung zur Version des Fragebogens erfolgte zufällig.

Die Datensätze der insgesamt 153 Versuchsteilnehmer wurden zusammengefasst und zu der im folgenden Abschnitt beschriebenen Prozedur der Itemreduktion verwendet.

## 2.3 Itemreduktion

Wir gehen davon aus, dass die wahrgenommene Attraktivität einer Software aus einer gewichteten Bewertung dieser Software bzgl. mehrerer Aspekte resultiert (siehe Hassenzahl 2001). Für den Fragebogen sollen zwei Arten von Items gefunden werden:

- Items, die die Attraktivität direkt messen,
- Items, die die Bewertung des Produkts auf den relevanten Aspekten messen.

Die 80 Items wurden deshalb in zwei Teilmengen aufgeteilt. Die erste Teilmenge enthielt alle Items (insgesamt 14), die Zustimmung/Ablehnung signalisieren, aber keine inhaltliche Bewertung vornahmen (Beispiele: gut/schlecht, unangenehm/angenehm, etc.). Die zweite Teilmenge enthielt die restlichen Items (insgesamt 66).

Die Faktorenanalyse der Zustimmungs/Ablehnungs Items ergab wie erwartet nur einen Faktor (sowohl nach dem Kaiser-Guttman-Kriterium als auch anhand des Scree-Tests, siehe Catell 1966). Dieser Faktor, den wir im folgenden als Attraktivität bezeichnen, erklärte 60% der aufgetretenen Varianz. Es wurden sechs Items aus dieser Teilmenge ausgewählt: *abstoßend / anziehend, unattraktiv / attraktiv, unangenehm / angenehm, unsympathisch / sympathisch, unerfreulich / erfreulich, schlecht / gut*.

Die Faktorenanalyse der zweiten Teilmenge ergab fünf Faktoren. Für die Ermittlung der Faktorenzahl wurde der Scree-Test angewendet (da das Kaiser-Guttman-Kriterium bei großen Variablenzahlen dazu tendiert, zu viele Faktoren zu extrahieren). Diese fünf Faktoren erklärten 53% der aufgetretenen Varianz. Die fünf Faktoren wurden nach den jeweils auf ihnen stark ladenden Items benannt als: *Durchschaubarkeit, Vorhersagbarkeit, Effizienz, Originalität* und *Stimulation*. Pro Faktor wurden vier Items gewählt, die auf dem jeweiligen Faktor besonders stark und auf den anderen Faktoren eher schwach luden.

Die restlichen Items wurden nun aus der Datenmatrix eliminiert. Die Daten wurden dann erneut mit der Faktorenanalyse untersucht. Hier ergaben sich erneut fünf Faktoren. Tabelle 2 zeigt die Ladung der ausgewählten Items auf diesen Faktoren. Die Items zur Attraktivität sind in der Tabelle nicht enthalten. Diese Items laden erwartungsgemäß hoch auf allen Faktoren.

Für die Erstellung des finalen Fragebogens wurden die Polung der verbliebenen 26 Items und deren Reihenfolge randomisiert. Der Fragebogen besteht also aus den Skalen Attraktivität (sechs Items), Durchschaubarkeit, Effizienz, Vorhersagbarkeit, Stimulation und Originalität (jeweils vier Items). Wir bezeichnen den Fragebogen im folgenden als *User Experience Fragebogen* (kurz UEQ).

Ein von Hassenzahl (2001) beschriebenes Rahmenmodell unterscheidet zwischen wahrgenommener ergonomischer Qualität, wahrgenommener hedonischer Qualität und der Attraktivitätsbeurteilung eines Produkts. Pragmatische und hedonische Qualität sind dabei Oberbegriffe, die verschiedene Qualitätsaspekte zusammenfassen. Pragmatische Qualität fokussiert dabei auf ziel- oder aufgabengerichtete Aspekte des Designs einer Software. Eine hohe pragmatische Qualität versetzt den User in die Lage seine Ziele effektiv und effizient zu erreichen. Hedonische Qualität fokussiert dagegen auf Qualitätsaspekte, die nicht primär aufgabenbezogen sind, z.B. Originalität. Das Attraktivitätsurteil ist eine globale Bewertung einer Software auf einer Zustimmungs-/Ablehnungsdimension, welches durch eine gewichteten Bewertung der einzelnen ergonomischen und hedonischen Qualitätsaspekte entsteht. Ordnet man die Dimensionen des UEQ in dieses Rahmenmodell ein, so sind Durchschaubarkeit, Effizienz und Vorhersagbarkeit pragmatische Qualitätsaspekte. Stimulation und Originalität können als hedonische Qualitätsaspekte aufgefasst werden.

Tabelle 2: Ladung der Items auf den Faktoren

Items	Faktor				
	Durchschaubarkeit	Originalität	Stimulation	Vorhersagbarkeit	Effizienz
<i>Verwirrend / Übersichtlich</i>	0,661				
<i>Schwer zu lernen / Leicht zu lernen</i>	0,856				
<i>Kompliziert / Einfach</i>	0,851				
<i>Unverständlich / Verständlich</i>	0,857				
<i>Herkömmlich / Neuartig</i>		0,849			
<i>Phantasielos / Kreativ</i>		0,785			
<i>Konservativ / Innovativ</i>		0,772			
<i>Konventionell / Originell</i>		0,79			
<i>Einschläfernd / Aktivierend</i>			0,601		
<i>Langweilig / Spannend</i>			0,661		
<i>Minderwertig / Wertvoll</i>			0,725	0,422	
<i>Uninteressant / Interessant</i>			0,838		
<i>Behindernd / Unterstützend</i>				0,505	
<i>Nicht Erwartungskonform / Erwartungskonform</i>	0,438			0,549	
<i>Unberechenbar / Voraussagbar</i>				0,791	
<i>Unsicher / Sicher</i>				0,74	
<i>Ineffizient / Effizient</i>					0,722
<i>Langsam / Schnell</i>					0,723
<i>Überladen / Aufgeräumt</i>					0,65
<i>Unpragmatisch / Pragmatisch</i>				0,419	0,635

### 3 Erste Ergebnisse zur Validierung

Bezüglich der Validierung des Fragebogens liegen bisher Ergebnisse aus zwei kleineren Usability Studien vor.

Die aufgabenorientierten Aspekte Durchschaubarkeit, Effizienz und Vorhersagbarkeit sollten stark negativ mit der Bearbeitungszeit einer Aufgabe korrelieren. Je schneller ein Benutzer seine Aufgaben erledigen kann, desto höher sollte er oder sie diese aufgabenorientierten Aspekte bewerten. Umgekehrt sollten die nicht-aufgabenorientierte Aspekte Stimulation und Originalität nicht oder nur gering mit der Bearbeitungszeit korrelieren.

Diese Hypothesen wurden im Rahmen eines Usability-Tests überprüft. 13 Testteilnehmer bearbeiteten dabei ein betriebswirtschaftliches Szenario und bewerteten das User Interface danach mit dem UEQ. Die Gesamtbearbeitungszeit variierte zwischen 33 und 65 Minuten.

Tabelle 3 zeigt die Korrelationen zwischen der Bearbeitungszeit der Aufgaben und den Bewertungen auf den Skalen des Fragebogens. Als Maß für die Reliabilität der Skalen wird noch der Alpha-Coeffizient angegeben.

Die Korrelationen zeigen das erwartete Muster. Durchschaubarkeit, Effizienz und Vorhersagbarkeit zeigen eine signifikante Korrelation ( $p < 0,05$ ) mit der Bearbeitungszeit. Stimulation und Originalität korrelieren nur schwach mit der Bearbeitungszeit. Die aufgestellten Hypothesen konnten damit bestätigt werden, was als ein erster Hinweis für die Konstruktvalidität des UEQ gewertet werden kann. Die gemessenen Alpha-Werte sind ein Hinweis auf eine ausreichende Reliabilität der Skalen, wobei hier die geringe Zahl der Teilnehmer berücksichtigt werden muss.

Tabelle 3: Korrelation der Skalen mit der Bearbeitungszeit und Cronbach's Alpha der Skalen.

Skala	Korrelation mit Bearbeitungszeit	Cronbach's Alpha
Attraktivität	-0,54	0,89
Durchschaubarkeit	-0,66*	0,82
Effizienz	-0,73*	0,73
Vorhersagbarkeit	-0,65*	0,65
Stimulation	0,10	0,76
Originalität	0,29	0,83

\* Signifikant mit  $p < 0,05$

In einer zweiten Studie wurden die Beziehungen der Skalen des UEQ zu den Skalen des AttrakDiff2 (Hassenzahl et al. 2003) untersucht. Dieser Fragebogen erlaubt die Messung der Qualitätsaspekte *Pragmatische Qualität*, *Hedonische Qualität* (die hier zusätzlich in die beiden Skalen *Identität* und *Stimulation* aufgeteilt ist) und *Attraktivität*. Die Attraktivitätskonzepte beider Fragebögen sind sehr ähnlich und sollten deshalb hoch miteinander korrelieren. Weiterhin sollten Durchschaubarkeit, Effizienz und Vorhersagbarkeit eine hohe Korrelation zur Skala Pragmatische Qualität im AttrakDiff2 zeigen. Stimulation und Originalität

sollten mit der Skala Stimulation des AttrakDiff2 hoch korrelieren. Bezüglich der Skala Identität des AttrakDiff2 können vorab keine Hypothesen formuliert werden.

Tabelle 4: Korrelation der Skalen des User Experience Fragebogens mit den Skalen des AttrakDiff2

		User Experience Fragebogen (UEQ)					
		Attraktivität	Durchsichtigkeit	Effizienz	Vorhersagbarkeit	Stimulation	Originalität
AttrakDiff2	Attraktivität	0,72 *	0,56 *	0,3	0,51 *	0,51 *	0,4
	Pragmatische Qualität	0,33	0,73 *	0,59 *	0,54 *	0,31	0,07
	Identität	0,45	0,45	0,29	0,62 *	0,3	0,32
	Stimulation	0,42	-0,17	-0,4	-0,14	0,72 *	0,64 *

\*Signifikant mit  $p < 0,05$

Diese Hypothesen wurden erneut im Rahmen eines Usability Tests untersucht. Hierbei bearbeiteten 16 Teilnehmer ein betriebswirtschaftliches Szenario. Unmittelbar nach der Bearbeitung des Szenarios beurteilte eine Hälfte der Teilnehmer die Benutzeroberfläche mit dem AttrakDiff2, die andere Hälfte mit dem UEQ. Danach wurde mit den Teilnehmern etwa 30 Minuten über die im Laufe der Aufgabenbearbeitung aufgetretenen Probleme diskutiert. Nach Abschluss der Diskussion beurteilte jeder Teilnehmer die Benutzeroberfläche erneut mit dem jeweils anderen Fragebogen. Pro Teilnehmer lagen also eine Bewertung der Benutzeroberfläche mit dem AttrakDiff2 und dem UEQ vor.

Tabelle 4 zeigt die Korrelationen der Skalen des User Experience Fragebogens mit den Skalen des AttrakDiff2. Die Ergebnisse zeigen weitgehend das erwartete Muster. Durchsichtigkeit und Vorhersagbarkeit korrelieren signifikant mit der Skala *Pragmatische Qualität* des AttrakDiff2. Die Skala Stimulation im AttrakDiff2 korreliert hoch mit den Skalen Originalität und Stimulation im UEQ. Die Skala Identität des AttrakDiff2 korreliert signifikant mit der Skala Vorhersagbarkeit des UEQ, aber nicht signifikant mit den Skalen Stimulation und Originalität.

## 4 Ausblick

Bei der Konstruktion des Benutzerfragebogens zur Messung der User Experience UEQ wurde durch die besondere Vorgehensweise bei der Itemauswahl darauf geachtet, möglichst viele relevante Produkteigenschaften zu berücksichtigen. Die gefundenen Faktoren bestätigen, dass ‚weichere‘ Kriterien, die eher der User Experience zuzuschreiben sind, für den Endnutzer eine ähnlich hohe Relevanz haben wie Kriterien der Usability im engeren Sinne (zwei Skalen vs. drei Skalen). Dies wird durch existierende Fragebögen nicht in dieser Weise abgedeckt. Die beschriebenen Validierungsuntersuchungen deuten auf angemessene Konstruktvalidität hin. Weitere hier nicht berichtete Studien zeigen außerdem, dass hypothesen-



konforme Unterschiede für verschiedene Benutzeroberflächen mit dem UEQ nachgewiesen werden können. Parallele Vergleichsdaten der deutschen und einer ersten englischen Version bestätigen zudem eine große Übereinstimmung der beiden Sprachversionen.

Der UEQ scheint ein hilfreiches und valides Messinstrument der User Experience zu sein, der andere Evaluationsmethoden gut ergänzen kann. Seine Qualitäten werden in weiteren Studien noch differenzierter zu untersuchen sein.

### Literaturverzeichnis

- Catell, R. B. (1966): The scree test for the number of factors. *Multivariate Behavioural Research*, Vol. 1, S. 245-276.
- DIN EN ISO 9241-10 (1996): Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten. Teil 10: Grundsätze der Dialoggestaltung. Berlin: Beuth Verlag.
- DIN EN ISO 9241-11 (1999): Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten. Teil 11: Anforderungen an die Gebrauchstauglichkeit – Leitsätze. Berlin: Beuth Verlag.
- Dzida, W.; Hofmann, B.; Freitag, R.; Redtenbacher, W.; Baggen, R.; Geis, T.; Beimel, J.; Zurheiden, C.; Hampe-Neteler, W.; Hartwig, R.; Peters, H. (2000): Gebrauchstauglichkeit von Software: ErgoNorm: Ein Verfahren zur Konformitätsprüfung von Software auf der Grundlage von DIN EN ISO 9241 Teile 10 und 11. Schriftenreihe der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin. Dortmund: Bundesanstalt für Arbeitsschutz und Arbeitsmedizin.
- Gediga, G.; Hamborg, K.-C. (1999): IsoMetrics: Ein Verfahren zur Evaluation von Software nach ISO 9241-10. In: H. Holling; G. Gediga (Hrsg.): *Evaluationsforschung*. Göttingen: Hogrefe. S. 195-234.
- Hamborg, K.-C. (2002): Gestaltungsunterstützende Evaluation von Software: Zur Effektivität und Effizienz des IsoMetricsL Verfahrens. In: Herczeg, W. Prinz; H. Oberquelle (Hrsg.): *Mensch & Computer 2002: Vom interaktiven Werkzeug zu kooperativen Arbeits- und Lernwelten*. Stuttgart: Teubner. S. 303-312.
- Hassenzahl, M. (2001): The effect of perceived hedonic quality on product appealingness. *International Journal of Human-Computer Interaction*, Vol. 13, Nr. 4, S. 481-499.
- Hassenzahl, M.; Burmester, M.; Koller, F. (2003): AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In: J.Ziegler; G. Szwillus (Hrsg.): *Mensch & Computer 2003. Interaktion in Bewegung*. Stuttgart: Teubner. S. 187-196.
- Kirakowski, J.; Corbett, M. (1993): SUMI: The Software Usability Measurement Inventory. *British Journal of Educational Technology*, Vol. 24, Nr. 3, S. 210–212.
- Laugwitz, B. (2001): Experimentelle Untersuchung von Regeln der Ästhetik von Farbkombinationen und von Effekten auf den Benutzer bei ihrer Anwendung im Benutzungsoberflächendesign. Berlin: dissertation.de – Verlag im Internet.
- Nielsen, J. (1994): Heuristic Evaluation. In: J. Nielsen; R.L. Mack (Hrsg.): *Usability Inspection Methods*. New York: Wiley. S. 25-62.
- Nielsen, J. (2001): Jakob Nielsen's Alertbox, August 5, 2001: First rule of usability: Don't listen to users. Available URL <http://www.useit.com/alertbox/20010805.html>.
- Preece, J.; Rogers, Y.; Sharpe, H. (2002): *Interaction design: Beyond human-computer interaction*. New York: Wiley.

Schrepp, M.; Held, T.; Laugwitz, B. (2004): Hedonische Aspekte betriebswirtschaftlicher Software. In: Keil-Slawik, R.; Selke, G. & Szwillus, G. (Hrsg.), Mensch & Computer 2004: Allgegenwärtige Interaktion. München: Oldenbourg. S. 127-136.

**Danksagung und Kontaktinformationen**

Wir danken Herrn Patrick Fischer für seine Unterstützung bei der statistischen Auswertung der Validierungsstudien.

Dr. Bettina Laugwitz / Dr. Martin Schrepp / Dr. Theo Held  
SAP AG, Dietmar-Hopp-Allee 16, 69190 Walldorf  
Email: [bettina.laugwitz@sap.com](mailto:bettina.laugwitz@sap.com) / [martin.schrepp@sap.com](mailto:martin.schrepp@sap.com) / [theo.held@sap.com](mailto:theo.held@sap.com)