

# Die Gewichtung von Usability-Aspekten anhand der „Analytic Hierarchy Process“ – Methode von Saaty

**Krisztin Pataki**  
TU Berlin  
Franklinstrasse 28/29  
10587 Berlin  
pataki@gp.tu-berlin.de  
www.ewi.tu-berlin.de

**Jochen Prümper**  
FHTW Berlin  
Treskowallee 8  
10318 Berlin  
j.pruemper@fhtw-berlin.de  
www.f3.fhtw-berlin.de

**Manfred Thüring**  
TU Berlin  
Franklinstrasse 28/29  
10587 Berlin  
thuering@gp.tu-berlin.de  
www.gp.tu-berlin.de

## Abstract

Im Rahmen der software-ergonomischen Evaluation spielt die Betrachtung der Wichtigkeit von Usability-Aspekten eine vernachlässigte Rolle. Wie in einer empirischen Studie, in der als unabhängige Variable drei verschiedene Nutzungskontexte und als abhängige Variablen die Beurteilung der software-ergonomischen Qualität, dimensionsspezifische Gewichte und Usability-Scores zum Einsatz kamen, konnte jedoch gezeigt werden, dass Nutzer diese Aspekte vor dem Hintergrund verschiedener

Nutzungskontexte sehr wohl als unterschiedlich wichtig wahrnehmen. Verfahren aus der Entscheidungstheorie können adaptiert werden, um systematisch Gewichte für Usability-Aspekte zu erheben. Diese Betrachtung von Gewichten in Verbindung mit software-ergonomischen Bewertungen liefert die Möglichkeit einer ganzheitlicheren Betrachtungsweise der ergonomischen Güte und erlaubt eine differenziertere Identifizierung von Entwicklungspotenzialen.

## Keywords

AHP, Gewichtung, ISONORM 9241/10, Online-Portale, Multiattribute Entscheidungen

## 1.0 Einleitung

Ziel partizipativer Softwaregestaltung ist die Anpassung von Software an die Bedürfnisse, Kenntnisse und Fertigkeiten des Nutzers unter Berücksichtigung des Nutzungskontextes. Die nutzer-zentrierte Evaluation stellt vor diesem Hintergrund ein wichtiges Instrument dar. Nutzer bewerten hierbei die software-ergonomische Qualität (Usability) der zu evaluierenden Software mittels entsprechender Fragebögen. Diese enthalten Bewertungsdimensionen, die globale Usability-Aspekte darstellen, die dann in Form von Aussagen (sog. Items) spezifiziert werden. Anhand von Rating-skalen bewerten Nutzer, in welchem Ausmaß diese Aussagen auf die zu evaluierende Software zutreffen. Es resultieren Nutzerurteile (in Form von Ratings), die die Bewertung der subjektiv wahrgenommen software-ergonomischen Qualität widerspiegeln.

Mittlerweile existieren eine Reihe fundierter Fragebögen zur Beurteilung der

software-ergonomischen Qualität (für einen Überblick siehe Sarodnick & Brau, 2006; Stanton et al. 2006). Allerdings sagen derartige Beurteilungen *per se* noch nichts über die Wichtigkeit einzelner Usability-Aspekte bezogen auf einen spezifischen Nutzungskontext aus.

Manche dieser Fragebögen – wie z.B. der ISONORM 9241/10 von Prümper (1997) oder der IsoMetrics von Hamburg und Gediga (2002) – sehen deshalb, wenn auch auf unterschiedliche Art und Weise, eine Gewichtung der jeweils zu beurteilenden software-ergonomischen Kriterien vor.

*Während z.B. in dem ISONORM 9241/10 von Prümper (1997) die sieben Grundsätze der Dialoggestaltung danach zu beurteilen sind, wie wichtig diese jeweils von den Nutzern für die Ausübung ihrer Tätigkeit beurteilt werden, werden in dem IsoMetrics von Hamburg und Gediga (2002) alle 75 Items, die zur Operationalisierung der DIN EN ISO 9241-10 (DIN EN ISO 9241-10 1993) dienen, danach beurteilt, wie wichtig die jewei-*

*ligen Aspekte für den Gesamteindruck der Software sind.*

Diese Gewichtung erscheint grundsätzlich interessant und plausibel, da vor dem Hintergrund verschiedener Nutzungskontexte und/oder Nutzergruppen verschiedene Usability-Aspekte möglicherweise eine unterschiedliche Bedeutung haben. So mag z.B. im Rahmen einer Beurteilung mittels eines Fragebogens, der die DIN EN ISO 9241-110 (DIN EN ISO 9241-110 2006) operationalisiert, dem Dialogprinzip „Individualisierbarkeit“ von Software eine hohe Bedeutung zugeschrieben werden, wenn sie von einer heterogenen Nutzergruppe verwendet wird, während der Aspekt der „Fehlertoleranz“ z.B. in einem besonders sicherheitsrelevanten Nutzungskontext bedeutend ist.

## 2.0 Das Problem verschiedener Bewertungsaspekte

Neben der Unterschiedlichkeit in der Operationalisierung der Wichtigkeit der Usability-Aspekte ist allen Beurteilungs-

verfahren gemeinsam, dass die Beurteilungen (Ratings) der software-ergonomischen Qualität und die Zuweisungen der Gewichte als getrennte Urteile verbleiben und nicht zu einem gemeinsamen Gesamtwert im Sinne eines „Usability-Scores“ verrechnet werden. Dem Durchführenden der jeweiligen Evaluation bleibt somit freigestellt, wie er mit diesen beiden quantitativen Bewertungsaspekten – also den Ratings auf der einen und den Gewichten auf der anderen Seite – verfährt.

## 2.1 Multiattribute Entscheidungsverfahren

Eine Lösung, Ratings und Gewichte zusammenzuführen – und damit gleichzeitig verschiedene Bewertungsmaße miteinander zu integrieren – bieten multiattribute Entscheidungsverfahren (Eisenführ & Weber 2003). Dabei handelt es sich um Methoden, die in komplexen Entscheidungssituationen, wenn nämlich eine Entscheidung zu Gunsten einer von verschiedenen Optionen vor dem Hintergrund mehrerer Entscheidungskriterien (sog. Attributen) erfolgen soll, Unterstützung anbieten.

Zum einen sehen diese Verfahren eine Gewichtung der Entscheidungskriterien vor. Zum anderen werden die Optionen, die zur Wahl stehen, hinsichtlich dieser Kriterien bewertet. Aus diesen Bewertungen ergeben sich für die Kriterien *Partialwerte* (auch genannt „Teilnutzenwerte“). Partialwerte und Gewichte werden dann miteinander multipliziert, so dass sich für jedes Kriterium *gewichtete Partialwerte* ergeben. Die Summierung der gewichteten Partialwerte über alle Kriterien führt zu *Gesamtwerten*, die sich für jede Option ergeben, so eine Rangreihe der Optionen gebildet werden kann. Eine Entscheidung soll dann zu Gunsten der Option mit dem größten Gesamtwert erfolgen. Diese Aggregationsmethode, also die Verknüpfung von Partialwerten und Gewichten, wird auch

als *MAU-Regel* (Multi-Attribute Utility) bezeichnet (Jungermann, Pfister und Fischer 2005).

Voraussetzung zur Anwendung der MAU-Regel ist, wie oben beschrieben, die Ermittlung von Gewichten. Eine populäre Methode, dies zu tun, ist die von Saaty (1990) entwickelte Methode *AHP* (Analytic Hierarchy Process; dt. Analytischer Hierarchieprozess) (s. auch Vargas 1990).

## 2.2 Das Gewichtungsverfahren AHP von Saaty (1990)

AHP (Analytic Hierarchy Process) von Saaty (1990) stellt ein mathematisch fundiertes Verfahren zur Gewichtung von Entscheidungskriterien im Kontext multiattributer Entscheidungssituationen dar. Im Rahmen von AHP werden zur Ermittlung der Wichtigkeit *w* alle Entscheidungskriterien, die für einen jeweiligen Kontext als relevant spezifiziert wurden, paarweise miteinander verglichen. Auf einer neunstufigen Skala (von 1 = „gleich wichtig“, 3 = „etwas wichtiger“, 5 = „sehr viel wichtiger“, 7 = „erheblich wichtiger“ bis 9 = „dominant wichtig“) wird dabei die subjektive Wichtigkeit der zu vergleichenden Kriterien ermittelt. Entscheidungsträger geben hierbei an, welche Wichtigkeit jeweils ein Kriterium im Vergleich zu einem anderen hat.

Abbildung 1 veranschaulicht anhand eines für Entscheidungssituationen typischen Beispiels „Standortwahl“ die Ermittlung der Gewichte für die Kriterien „Verkehrsanbindung“, „Freizeitangebot“ und „Grundstückspreis“. Aus den Paarvergleichen ergibt sich die in Abbildung 1 dargestellte Paarvergleichsmatrix. Die Ermittlung der Gewichte für jedes Kriterium erfolgt, indem in einem ersten Schritt die Zeilenelemente der Paarvergleichsmatrix durch die Summe ihrer Spalten dividiert werden.

In unserem Beispiel bedeutet dies, dass das Zeilenelement „2“ für den Vergleich von „Verkehrsanbindung“ und „Freizeitangebot“ dividiert wird durch die Spaltensumme „8“. Das Ergebnis lautet entsprechend  $2/8=0,25$ .

In einem zweiten Schritt werden dann diese, durch die Spaltensumme normalisierten, Zeilenelemente jeder Zeile der Matrix gemittelt.

| Welches Kriterium ist Ihnen für das Ziel „Standortwahl“ wichtiger und in welchem Ausmaß? |                    |                  |                   |
|--|--------------------|------------------|-------------------|
|  | Verkehrs-anbindung | Freizeit-angebot | Grund-stückspreis |
| Verkehrs-anbindung   | 1                  | 2                | 3                 |
| Freizeit-angebot   | 1/2                | 1                | 1/5               |
| Grund-stückspreis  | 1/3                | 5                | 1                 |

• Verkehrsanbindung wird im Vergleich zu Freizeitangebot zwischen „gleich wichtig“ und „etwas wichtiger“ bewertet.  
 • Verkehrsanbindung wird im Vergleich zu Grundstückspreis als „etwas wichtiger“ bewertet und  
 • Freizeitangebot wird im Vergleich zu Grundstückspreis als „sehr viel weniger wichtiger“ bewertet.

Abb 1: Ermittlung der Wichtigkeiten am Beispiel „Standortwahl“<sup>1</sup>

In unserem Beispiel erhält damit die Zeile „Verkehrsanbindung“ das Gewicht  $w = 0,2$  gemäß der Berechnung  $1/(1+0,5+0,33) + 2/(2+1+5) + 3/(3+0,2+1)$ . Entsprechend wird mit der Zeile „Freizeitangebot“ ( $w = 0,1$ ) und „Grundstückspreis“ ( $w = 0,7$ ) verfahren.

Die Summe der so resultierenden, normalisierten Kriteriumsgewichte ergibt stets eins.

Wie oben beschrieben, können nun die mittels AHP erhobenen Gewichte entsprechend der MAU-Regel mit den Partialwerten zu Gesamtwerten verknüpft werden.

<sup>1</sup> AHP trifft eine Annahme hinsichtlich der *Reziprozität* von Vergleichen. Wird z.B. das Kriterium „Grundstückspreis“ als „etwas wichtiger“ als das Kriterium „Verkehrsanbindung“ bewertet (Skalenwert 3) so ergeben sich umgekehrte Vergleichswerte für das Kriterium „Grundstückspreis“ gegenüber „Standortwahl“ (Skalenwert 1/3). Daher stellen die Werte der Paarvergleichsmatrix unterhalb der Diagonalen umgekehrte Werte derer oberhalb der Diagonalen dar.

In unserem Beispiel „Standortwahl“ muss eine Entscheidung zu Gunsten eines Standortes getroffen werden. Mehrere Standorte kommen als Optionen in Frage. Alle Standorte besitzen auf den Kriterien „Verkehrsanbindung“, „Freizeitangebot“ und „Grundstückspreis“ entsprechende Ausprägungen, die im Vorfeld definiert werden. So werden beispielsweise Preise für baureifes Land von über 250 €/m<sup>2</sup> als sehr hoch mit der Bewertung 5 oder sehr niedrige Grundstückspreise unter 50 €/m<sup>2</sup> als sehr niedrig mit der Bewertung 1 belegt. Die Bewertungen für jedes Kriterium, die sogenannten Partialwerte, werden schließlich mit den zugehörigen Gewichten multipliziert. Erhält beispielsweise Standort A für das Kriterium „Grundstückspreis“ den Partialwert  $pw = 2$  und wird mit dem zugehörigen Kriteriumsgewicht von  $w = 0.7$  multipliziert, so ergibt sich ein gewichteter Partialwert von  $pw_g = 1.4$ . Gewichtete Partialwerte werden dann für alle Kriterien ermittelt und aufaddiert. Verfährt man derart für alle Standorte und Kriterien, resultieren Gesamtwerte, die anzeigen, welcher Standort zu wählen ist – nämlich der mit dem höchsten Nutzen. Entsprechend werden diese Gesamtwerte auch als „Nutzwerte“ bezeichnet.

### 3.0 AHP im Kontext software-ergonomischer Evaluation

Um im Rahmen der software-ergonomischen Evaluation Gewichte zu ermitteln, kann die AHP Methode mit einem Fragebogen zur Erhebung der software-ergonomischen Qualität verknüpft werden.

Zur Ermittlung von Gewichten der in Fragebögen enthaltenen Usability-Aspekte sieht AHP in einem ersten Schritt die Gewichtung der Items und in einem zweiten Schritt die der Dimensionen vor. Hierzu werden zunächst alle Items einer Bewertungsdimension paarweise miteinander verglichen und ihre Wichtigkeit in Bezug auf ihre übergeordnete Bewertungsdimension bestimmt. So verfährt man mit allen Items jeder Bewertungsdimension.

Die Gewichte, die sich für die Usability-Kriterien auf Itemebene ergeben, wer-

den als itemspezifische Gewichte  $w_i$  bezeichnet. Nach der Ermittlung itemspezifischer Gewichte wird die Gewichtung entsprechend der AHP-Methode für die Kriterien auf Dimensionsebene fortgesetzt. Das Ergebnis sind dimensionsspezifische Gewichte  $w_d$ .

Item- und dimensionsspezifische Gewichte werden dann zu Gesamtgewichten  $w_g$  zusammengeführt. Hierbei wird jedes itemspezifische Gewicht  $w_i$  mit dem dazugehörigen dimensionsspezifischen Gewicht  $w_d$  multipliziert (vgl. Saaty, 1990). Es resultieren Größen, die die Relevanz der Aspekte auf Item- und Dimensionsebene enthalten. Eine derartige Gewichtung wird als hierarchische Gewichtung bezeichnet (Pöyhönen & Hämäläinen 1998), da sie alle Kriterien des Fragebogens betrachtet (für ein Beispiel siehe Abbildung 2).

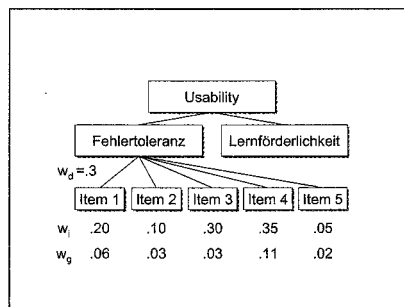


Abb 2: Gewichte für die Dimension „Fehlertoleranz“ und ihre zugehörigen Items

Abbildung 2 veranschaulicht die Antwort auf die Frage, wie wichtig – in einem bestimmten Nutzungskontext – das Dialogprinzip „Fehlertoleranz“ gegenüber dem der „Lernförderlichkeit“ eingeschätzt wird. Daraus resultieren die Gewichte für die Dimensionen „Fehlertoleranz“ und „Lernförderlichkeit“. Diese werden dann mit den itemspezifischen Gewichten verrechnet. So ergeben sich für die Items der Dimension „Fehlertoleranz“ die in Abbildung 2 dargestellten, itemspezifischen Gewichte  $w_i$ , die dann mit dem dimensionsspezifischen Ge-

wicht  $w_d$  mittels Multiplikation die Gesamtgewichte  $w_g$  ergeben.

Gesamtgewichte können nun mit den zuvor ermittelten Ratings zu Usability-Scores verrechnet werden, indem jedes auf Itemebene entstandene Gesamtgewicht  $w_g$  mit dem ihm zugehörigen Rating  $R$  multipliziert wird, so dass gewichtete Ratings  $R_g$  entstehen. Gewichtete Ratings  $R_g$  werden für jede Dimension addiert. Als Ergebnis resultieren dimensionsspezifische Usability-Scores  $US_d$ , die in der Summe Nutzenwerte in Form von dimensionslosen Usability-Scores  $US$  ergeben (für ein Beispiel siehe Abbildung 3 und 4).

Für eine Software ergeben sich aus den Ratings  $R$  für jedes Item durch Multiplikation der Gesamtgewichte  $w_g$  gewichtete Ratings  $R_g$ . Beispielfhaft resultiert für die Dimension „Fehlertoleranz“ aus der Summe der gewichteten Ratings für alle Items 1 bis 5 der dimensionsspezifische Usability-Score  $US_d = .93$  (vgl. Abbildung 3).

| Dimension „Fehlertoleranz“ |   |       |       |        |
|----------------------------|---|-------|-------|--------|
|                            | R | $w_g$ | $R_g$ | $US_d$ |
| Item 1                     | 3 | .06   | .18   | .93    |
| Item 2                     | 2 | .03   | .06   |        |
| Item 3                     | 3 | .09   | .27   |        |
| Item 4                     | 2 | .20   | .40   |        |
| Item 5                     | 1 | .02   | .02   |        |

Abb 3: Berechnung von Usability-Scores für die Dimension „Fehlertoleranz“

Die Addition der dimensionsspezifischen Usability-Scores  $US$  für alle im Fragebogen enthaltenen Dimensionen ergibt für das Beispiel einen Usability-Score von  $US = 4.21$ .

|                              | US <sub>d</sub> |
|------------------------------|-----------------|
| Aufgabenangemessenheit       | .86             |
| Selbstbeschreibungsfähigkeit | .75             |
| Steuerbarkeit                | .52             |
| Erwartungskonformität        | .54             |
| <b>Fehlertoleranz</b>        | <b>.93</b>      |
| Individualisierbarkeit       | .47             |
| Lernförderlichkeit           | .68             |
| <b>Usability-Score (US)</b>  | <b>4.21</b>     |

Abb 4: Aggregation von Usability-Scores

#### 4.0 Eine Studie zur Gewichtung von Usability-Aspekten

In einer Studie zur Evaluation der software-ergonomischen Qualität von Websites aus verschiedenen Nutzungskontexten wurde das Gewichtungsverfahren AHP eingesetzt. Dabei standen zwei Fragen im Vordergrund. Die erste Frage lautete: „Inwiefern werden Usability-Aspekte von Nutzern als unterschiedlich wichtig wahrgenommen?“. Die zweite Frage war: „Inwiefern verändern Usability-Scores den Gesamtnutzen von Systemen?“

#### 4.1 Methode

Bei der vorliegenden Studie handelt es sich um ein Laborexperiment, in dem Probanden zur software-ergonomischen Qualität von Websites aus drei unterschiedlichen Nutzungskontexten sowie zur Wichtigkeit der in diesem Zusammenhang beurteilten Usability-Aspekte befragt wurden.

Die unabhängige Variable „Nutzungskontext“ hatte die drei Ausprägungen „Online-Banking“, „Online-Reisereservierung“ und „Online-Auktion“. Als abhängige Variablen dienten die Beurteilung der *software-ergonomischen Qualität* in Form von Ratings, *dimensionsspezifische Gewichte* und *Usability-Scores*.

#### 4.1.1 Versuchsmaterial

**Nutzungskontext:** Da die Probanden sich im Rahmen der Untersuchung für eine Website aus demjenigen der drei vorgegebenen Nutzungskontexte entscheiden sollten, aus dem sie die meiste Erfahrung mitbrachten, hatten sie innerhalb der vorgegebenen drei Nutzungskontexte „freie Wahl“ bei der Entscheidung für eine bestimmte Site. Für den Nutzungskontext „Online-Banking“ wählte der Großteil der Probanden vier verschiedene Bankportale; für den Nutzungskontext „Online-Reisereservierung“ vier verschiedene Portale zur Flugbuchung und eines zur Hotelbuchung und für den Nutzungskontext „Online-Auktion“ eine bekannte Internet-Auktionsplattform.

**Prüfaufgaben:** Die Untersuchung war nicht an vorgegebene Prüfaufgaben gekoppelt, sondern griff gezielt die Vorerfahrungen der Probanden auf, die diese bereits in einem spezifischen Nutzungskontext gesammelt hatten.

**ISONORM 9241/10:** Zur Bewertung der software-ergonomischen Qualität wurde der Fragebogen ISONORM 9241/10 von Prümper (1997) herangezogen. In Anlehnung an die Dialogprinzipien der DIN EN ISO 9241-10 (1993) ermittelt dieses Instrument anhand sieben Bewertungsdimensionen, die mit je fünf Items (Aussagen) spezifiziert werden, die software-ergonomische Qualität auf einer bipolaren, siebenstufigen Likert-Skala von -3 (- - -) bis +3 (+++).

**Bewertungssoftware ExpertChoice®:** Sowohl die ISONORM 9241/10-Bewertung als auch die AHP-Gewichtungen erfolgten computergestützt mittels der Bewertungssoftware ExpertChoice® (Version 11.3) (vgl. Meixner & Haas 2002).

#### 4.1.2 Versuchspersonen

An der Untersuchung nahmen N = 90 Studierende (45 Studentinnen und 45 Studenten) der Berliner Universitäten im Einzelversuch teil. Jeder der drei Nutzungskontexte repräsentierte eine Untersuchungsgruppe und war mit jeweils 30 Studierenden gleich groß besetzt. Männliche und weibliche Teilnehmer waren über alle drei Untersuchungsgruppen gleich verteilt. Das Durchschnittsalter betrug M = 28,5 Jahre.

#### 4.1.3 Versuchsablauf

In einem ersten Schritt suchten sich die Probanden aus dem Umfeld der drei vorgegebenen Nutzungskontexte die Website aus, mit der sie in der Vergangenheit die meiste Erfahrung gesammelt hatten. In einem zweiten Schritt wurden die Untersuchungsteilnehmer gebeten, die Tätigkeiten, die sie im Allgemeinen mit dieser Website ausführten, schriftlich zu fixieren und sich während der anschließenden Beurteilung der software-ergonomischen Qualität und der Wichtigkeit der Usability-Aspekte an dieser selbstbezogenen Tätigkeitsanalyse zu orientieren.

Im Anschluss erfolgte die Bewertung der software-ergonomischen Qualität mittels des ISONORM 9241/10 sowie die Erhebung der Gewichte mittels der AHP-Methode. Neben der Beurteilung der 35 Items des ISONORM 9241/10 nahmen die Probanden bei sieben Dimensionen á fünf Items 70 Paarvergleiche auf Itemebene zzgl. 21 Paarvergleiche auf Dimensionsebene vor. Ein Versuch dauerte inkl. Instruktion pro Person im Durchschnitt 30 Minuten.

#### 4.2 Ergebnisse

##### 4.2.1 Beurteilung der Software-Ergonomie

Abbildung 5 gibt einen Überblick über die software-ergonomische Qualität der in den drei jeweiligen Nutzungskontexten

texten zur Beurteilung herangezogenen Portale. Dabei zeigt sich, dass die in den drei Nutzungskontexten beurteilten Portale im Durchschnitt über alle sieben Dialogprinzipien mit einem ISONORM 9241/10-Gesamtwert zwischen 0.5 und 0.9 lediglich eine mittelmäßig ergonomische Qualität zugeschrieben wird. Online-Portale der drei Nutzungskontexte unterscheiden sich signifikant hinsichtlich der Aspekte „Selbstbeschreibungsfähigkeit“ und „Steuerbarkeit“. Im Vergleich zu Portalen aus den Bereichen „Reisereservierung“ und „Auktion“ wurden Portale aus dem Bereich „Banking“ als weniger selbstbeschreibungsfähig bewertet. Online-Portale vor dem Hintergrund des Nutzungskontextes „Auktion“ wurden vergleichend zu Portalen aus den Bereichen „Banking“ und „Reisereservierung“ als steuerbarer beurteilt (vgl. Abbildung 5).

|   | Banking | Reisereservierung | Auktion | F     | p <  |
|---|---------|-------------------|---------|-------|------|
| Aufgabenangemessenheit                  | .87     | .72               | 1.10    | 0.793 | n.s. |
| Selbstbeschreibungsfähigkeit            | .21     | .56               | .79     | 3.185 | .05  |
| Steuerbarkeit                           | .12     | .23               | .69     | 2.880 | .05  |
| Erwartungskonformität                   | 1.32    | .80               | 1.29    | 1.561 | n.s. |
| Fehlertoleranz                          | .79     | .70               | .77     | 0.059 | n.s. |
| Individualisierbarkeit                  | -.59    | -.31              | -.15    | 1.056 | n.s. |
| Lernförderlichkeit                      | 1.36    | 1.15              | 1.56    | 1.049 | n.s. |
| ISONORM-Gesamt (ISO <sub>gesamt</sub> ) | .58     | .55               | 0.86    | 1.223 | n.s. |

Abb 5: Beurteilung der Software-Ergonomie anhand des ISONORM 9241/10

#### 4.2.2 Gewichte

Nutzer nehmen Usability-Kriterien in Abhängigkeit der Nutzungskontexte als verschieden wichtig war. Dies lässt sich sowohl für dimensions- als auch für itemspezifische Aspekte der software-ergonomischen Beurteilung zeigen. Die folgenden Ausführungen beschränken sich auf die dimensionsspezifische Gewichtung.

Dem Aspekt „Lernförderlichkeit“ wurde für den Nutzungskontext „Auktion“ eine

signifikant höhere Relevanz zugeschrieben als für die Nutzungskontexte „Banking“ und „Reisereservierung“ (vgl. Abbildung 6). Des Weiteren beurteilten Nutzer das Dialogprinzip „Selbstbeschreibungsfähigkeit“ als signifikant wichtiger für den Anwendungsbereich „Reisereservierung“ als für die Bereiche „Banking“ und „Auktion“. Zudem wurde für alle der drei Nutzungskontexte der Aspekt „Steuerbarkeit“ als am wenigsten bedeutsam im Vergleich zu den verbleibenden Gestaltungsprinzipien bewertet. Dem Aspekt der „Lernförderlichkeit“ wurde in dem Nutzungskontext „Auktion“ im Vergleich zu den verbleibenden Usability-Aspekten die größte Bedeutsamkeit zugewiesen (vgl. Abbildung 6).

|                              | Banking | Reisereservierung | Auktion | F     | p <  |
|------------------------------|---------|-------------------|---------|-------|------|
| Aufgabenangemessenheit       | .17     | .17               | .14     | .548  | n.s. |
| Selbstbeschreibungsfähigkeit | .14     | .18               | .14     | 5.587 | .05  |
| Steuerbarkeit                | .09     | .07               | .06     | 2.000 | n.s. |
| Erwartungskonformität        | .15     | .17               | .15     | 0.764 | n.s. |
| Fehlertoleranz               | .20     | .17               | .19     | .449  | n.s. |
| Individualisierbarkeit       | .10     | .09               | .12     | 1.920 | n.s. |
| Lernförderlichkeit           | .16     | .16               | .21     | 5.656 | .05  |

Abb 6: Mittlere dimensionsspezifische Gewichte für die sieben Dialogprinzipien der DIN EN ISO 9241-110 in Abhängigkeit vom Nutzungskontext

#### 4.2.3 Usability-Scores

Die Betrachtung der Usability-Scores – also die Verrechnung von Ratings und Gewichten (vgl. Absatz 3.2) – führt im vorliegenden Kontext nicht zu einer Veränderung der Gesamtbewertung der Online-Portale hinsichtlich ihrer software-ergonomischen Qualität. Usability-Scores und ISONORM 9241/10-Gesamtwerte weisen in dieselbe Richtung (vgl. Abbildung 5 und 7).

|                              | Banking | Reisereservierung | Auktion |
|------------------------------|---------|-------------------|---------|
| Aufgabenangemessenheit       | .15     | .14               | .17     |
| Selbstbeschreibungsfähigkeit | .04     | .10               | .11     |
| Steuerbarkeit                | .03     | .03               | .05     |
| Erwartungskonformität        | .19     | .14               | .20     |
| Fehlertoleranz               | .16     | .12               | .15     |
| Individualisierbarkeit       | -.06    | -.01              | -.01    |
| Lernförderlichkeit           | .27     | .41               | .41     |
| Usability-Score (US)         | .78     | .75               | 1.08    |

Abb 7: Dimensionsspezifische Usability-Scores und Usability-Scores für die untersuchten Nutzungskontexte

## 5.0 Diskussion

Die vorliegende Untersuchung zeigte, dass sich die Online-Portale der drei untersuchten Anwendungsbereiche „Online-Banking“, „Online-Reisereservierung“ und „Online-Auktion“ hinsichtlich ihrer software-ergonomischen Qualität für einige Usability-Aspekte signifikant voneinander unterscheiden. Des Weiteren wurde verschiedenen Kriterien zur Bewertung der software-ergonomischen Qualität in Abhängigkeit von dem jeweiligen Nutzungskontext eine unterschiedliche Wichtigkeit beigemessen. Darüber hinaus zeigte sich, dass Dialogprinzipien, die eine höhere software-ergonomische Beurteilung erfahren, ebenfalls eine höhere Wichtigkeit erhalten – und *vice versa*. Vor diesem Hintergrund weisen ISONORM 9241/10-Gesamtwerte und Usability-Scores in die gleiche Richtung.

Usability-Scores würden sich von der Gesamtbewertung der software-ergonomischen Qualität dann unterscheiden, wenn z.B. hinsichtlich der software-ergonomischen Qualität niedrige Beurteilungen aber hohe Gewichtungen für Usability-Aspekte resultieren, d.h. software-ergonomische Beurteilungen und Wichtigkeiten in unterschiedliche Richtungen weisen. Dieser Zusammenhang ist durchaus denkbar, trat jedoch in dieser Untersuchung nicht auf.

Trotz der überzeugenden Vorteile, die die Anwendung der AHP-Methode in Verbindung mit dem Einsatz des ISO-NORM 9241/10 im Rahmen der Gewichtung von Usability-Aspekten mit sich bringt, mag ihre Anwendung in der betrieblichen Usability-Praxis an zeitökonomische Grenzen stoßen. Möchte man zur Gewichtung von Usability-Aspekten eine effizientere Art der Anwendung der AHP-Methode realisieren, empfiehlt sich die Verknüpfung der Methode mit dem Kurzfragebogen zur Software-Evaluation ISONORM 9241/110-S (siehe Pataki et al. 2006). Dieses Instrument verfügt ebenfalls über die bisher beschriebenen sieben Bewertungsdimensionen, operationalisiert diese jedoch lediglich mit je drei, anstatt fünf Items. Durch den Einsatz des ISONORM 9241/110-S reduziert sich zum einen der Aufwand software-ergonomischer Beurteilung (21 Items gegenüber zuvor 35 Items) und zum anderen der der Gewichtung deutlich, da nun nur noch 21 Paarvergleiche auf Itemebene (gegenüber zuvor 70 Paarvergleichen) und 21 Paarvergleiche auf Dimensionsebene durchgeführt werden müssen.

Vor dem Hintergrund einer zeitlich ökonomischen Evaluation in der Praxis ist ebenfalls zu diskutieren, dass die Ermittlung von Wichtigkeiten von Usability-Aspekten die Möglichkeit liefert, Gewichte als domänenspezifisch zu beschreiben, falls diese z.B. über eine große Stichprobe erhoben wurden. Diese bilden die Bedeutung von Usability-Aspekten vor dem Hintergrund verschiedener Nutzungskontexte ab. Mittels Sensitivitätsanalysen (vgl. Eisenführ & Weber 2003) werden diese Gewichte „nachgebessert“. So können beispielsweise die in der vorliegenden Untersuchung erhobenen Gewichte für die Domäne „Online-Anwendungen“ in nachfolgenden Evaluationen zu Online-Portalen aus den untersuchten Bereichen eingesetzt werden, ohne dass Nutzer diese neu ermit-

teln müssen. Bewertungen der software-ergonomischen Qualität in Form von Ratings werden dann mit vorab ermittelten domänenspezifischen Gewichten zu Usability-Scores verknüpft. Nutzer müssen dann lediglich nur noch mittels Sensitivitätsanalysen die Gewichte justieren, falls diese nicht den subjektiv wahrgenommen Bedeutungen für die jeweiligen Usability-Aspekte entsprechen.

Die vorliegenden Überlegungen und ersten empirischen Untersuchungen zur Gewichtung von Usability-Aspekten im Rahmen der Nutzerevaluation zeigen, dass diese für verschiedene Nutzungskontexte als unterschiedlich relevant wahrgenommen werden. Aufgrund der unterschiedlichen Gewichtung von Usability-Aspekten sollten die entsprechenden Wichtigkeiten in die Betrachtung der software-ergonomischen Qualität miteinbezogen werden. Dies ermöglicht eine ganzheitlichere Betrachtungsweise der software-ergonomischen Qualität. In einer differenzierteren Art und Weise werden Entwicklungspotenziale aufgezeigt. Wird eine Software hinsichtlich verschiedener Usability-Aspekte als wenig benutzungsfreundlich beurteilt, diesen Aspekten aber eine hohe Gewichtung verliehen, dann sollten diese Aspekte in der Weiterentwicklung eher berücksichtigt werden als die, denen eine niedrigere Wichtigkeit zugeschrieben wird.

## 6.0 Literaturverzeichnis

DIN EN ISO 9241-10 (1993): Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten - Teil 10: Grundsätze der Dialoggestaltung. Berlin: Beuth.

DIN EN ISO 9241-110 (2006): Ergonomie der Mensch-System-Interaktion - Teil 110: Grundsätze der Dialoggestaltung. Berlin: Beuth.

Eisenführ, F.; Weber, M. (2002): Rationales Entscheiden. Berlin: Springer.

Jungermann, H.; Pfister, R.; Fischer, K. (2005): Einführung in die Entscheidungspsychologie. Berlin: Springer.

Hamborg, K.C.; Gediga, G. (2002). IsoMetrics<sup>L</sup> – Fragebogen zur Evaluation von grafischen Benutzungsschnittstellen. Osnabrück

Meixner, O.; Haas, R. (2002): Computergestützte Entscheidungsfindung. ExpertChoice und AHP - innovative Werkzeuge zur Lösung komplexer Probleme. Heidelberg: Redline Wirtschaft.

Pataki, K.; Sachse, K.; Prümper, J.; Thüring, M. (2006): ISONORM 9241/10-S: Kurzfragebogen zur Software-Evaluation. In Lösel, F. (Hrsg.): Berichte über den 45. Kongress der Deutschen Gesellschaft für Psychologie. Lengerich: Pabst Science Publishers, S. 258-259.

Pöyhönen, M.; Hämäläinen, R. (1998): Notes on Weighing Biases in Value Trees. Journal of Behavioral Decision Making, Vol. 11, Nr. 2, S. 139–150.

Prümper, J. (1997): Der Benutzungsfragebogen ISONORM 9241/10: Ergebnisse zur Reliabilität und Validität. In: Liskowsky, R.; Velichkovsky, B.M.; W. Wüschmann (Hrsg.): Software-Ergonomie '97 - Usability Engineering: Integration von Mensch-Computer-Interaktion und Software-Entwicklung. Stuttgart: Teubner, (S. 253-262).

Saaty, L. (1990): How to make a decision: The Analytic Hierarchy Process. European Journal of Operational Research, Vol. 48, S. 9-26.

Sarodnick, F.; Brau, H. (2006): Methoden der Usability Evaluation. Wissenschaftliche Grundlagen und praktische Anwendung. Bern: Huber.

Stanton, N. A.; Salmon, P. M.; Walker, G. H.; Baber, Ch.; Jenkins, D.P. (2006): Human Factor Methods. A Practical Guide for Engineering and Design. Hampshire: Ashgate.

Vargas, L. (1990): An Overview of the Analytic Hierarchy Process and its Applications, European Journal of Operational Research Vol. 48, Nr. 1, S. 2-8