

Mit dem AttrakDiff die Attraktivität interaktiver Produkte messen

Marc Hassenzahl
Technische Universität
Darmstadt
Institut für Psychologie
Steubenplatz 12, 64293 Darmstadt
hassenzahl@
psychologie.tu-darmstadt.de

Abstract

Das AttrakDiff ist ein semantisches Differential zur Messung der pragmatischen Qualität, hedonischen Qualität und der Attraktivität interaktiver Produkte. Der vorliegende Beitrag erläutert zunächst diese Konstrukte. Danach wird auf im Zusammenhang mit semantischen

Differentialen häufig gestellte Fragen und den spezifischen Eigenschaften des AttrakDiff im Vergleich zu anderen, im Rahmen des usability engineerings üblichen, Fragebögen eingegangen. Den Abschluss bilden Überlegungen zur Einbindung des AttrakDiffs in eigene Projekte.

Keywords

Evaluation; Fragebogen; Attraktivität; hedonische Qualität

1.0 Einleitung

Das usability engineering beschäftigt sich mit der systematischen Sicherung der Gebrauchstauglichkeit interaktiver Produkte. Gebrauchstauglichkeit (usability) wird dabei als ein Qualitätsziel verstanden, das durch einen iterativen Prozess der Analyse, Gestaltung und Überprüfung erreicht werden kann (vgl. 3). Dabei kommt der empirischen, möglichst objektiven Überprüfung der Zielerreichung am Ende eines Zyklus eine wichtige Rolle zu.

Übliche Werkzeuge zur Überprüfung interaktiver Produkte im Rahmen des usability engineering sind Fragebögen (siehe auch Hamborg in diesem Band). Der vorliegende Beitrag beschäftigt sich mit dem Attrak-Diff, einem Fragebogen zur Messung der wahrgenommenen pragmatischen Qualität, hedonischen Qualität und der Attraktivität eines interaktiven Produkts (siehe 5).

2.0 Das attrakdiff und seine Konstrukte

Das Attrakdiff gehört zur Klasse der semantischen Differentiale. Es besteht aus 28 bipolaren, siebenstufigen Items (siehe Abbildung 1 für Beispielitems). Diese 28 einzelnen Items werden zu vier Skalen zusammengefasst. Sie operationalisieren die folgenden Konstrukte (für eine ausführliche Beschreibung des zugrunde liegenden Modells siehe 4

hässlich	☐	☐	☐	☐	☐	☐	☐	schön
praktisch	☐	☐	☐	☐	☐	☐	☐	unpraktisch
sympathisch	☐	☐	☐	☐	☐	☐	☐	unsympathisch
umständlich	☐	☐	☐	☐	☐	☐	☐	direkt
stilvoll	☐	☐	☐	☐	☐	☐	☐	stillos
voraussagbar	☐	☐	☐	☐	☐	☐	☐	unberechenbar

Abbildung 1: Einige Items des AttrakDiff

Pragmatische Qualität (PQ):
Die Fähigkeit eines Produkts, das Bedürfnis nach Zielerreichung zu befriedigen, indem es nützliche und benutzbare Funktionen bereit stellt. Typische Produktattribute sind: praktisch, voraussagbar, übersichtlich, handhabbar.

Hedonische Qualität – Stimulation (HQS):

Die Fähigkeit eines Produkts, das Bedürfnis nach Verbesserung der eigenen Kenntnisse und Fertigkeiten zu befriedigen. Typische Produktattribute sind: fesselnd, kreativ, originell, herausfordernd.

Hedonische Qualität – Identität (HQS):

Die Fähigkeit eines Produkts, relevanten Anderen selbstwertdienliche Botschaften zu kommunizieren. Typische Produktattribute sind: bringt mich den Leuten näher, fachmännisch, verbindend, stilvoll.

Attraktivität (ATT):

Globale positiv-negativ Bewertung des Produkts: gut, attraktiv, angenehm.

PQ entspricht im weitesten Sinne der Gebrauchstauglichkeit. Die übrigen Konstrukte hingegen, stellen zusätzliche, beim usability engineering bisher wenig beachtete, Qualitätsaspekte dar.

3.0 Häufige Fragen zum semantischen Differential

Das AttrakDiff ist ein kontextspezifisches semantisches Differential, das speziell für die Bewertung interaktiver Produkte entworfen wurde. Semantische Differentiale gehören zum Standardinstrumentarium der Psychologie. Bortz und Döring 1 weisen allerdings auf drei problematische Aspekte hin:

Unterschiede im Verständnis einzelner Items zwischen Beurteilern:

U.U. werden einzelne Items von Beurteilern unterschiedlich aufgefasst. Das AttrakDiff versucht dies durch die recht hohe Zahl von sieben Items pro Konstrukt zu kompensieren. Daraus folgt aber auch, dass bei der Interpretation vornehmlich dem Skalenwert (also dem Mittelwert der zu einer Skala gehörenden Items) und nicht so sehr den einzelnen Items Bedeutung zugewiesen werden sollte.

Unterschiede in der Bedeutung einzelner Items in Abhängigkeit der untersuchten Produkte.

Je nach Produkt können einzelne Items unterschiedliche Bedeutung annehmen. Dies ist allerdings nur dann problematisch, wenn sehr unterschiedliche Produkte verglichen werden sollen (z.B. Websites mit Mobiltelefonen). Da dies im Rahmen der Produktentwicklung nur selten der Fall sein dürfte, ist dieses allgemeine Problem des semantischen Differentials nicht so dringlich. Allerdings sollte man sich als Anwender des Fragebogens im Klaren darüber sein, dass die Konstrukte ihre Bedeutung je nach Produktkategorie zumindest leicht verändern können.

Akzeptanzprobleme bei unvorbereiteten Urteilern.

Semantische Differentiale stoßen häufig auf Akzeptanzprobleme, da einige Urteile

ungewohnt sind: »Bringt mich« ein Bankautomat »den Leuten näher« oder »trennt« er »mich von den Leuten«?. Es sind allerdings auch Produkte denkbar, bei denen die Beurteilung dieses Items weniger ungewohnt wäre, z.B. Mobiltelefone oder persönliche Websites. Der Wunsch, das AttrakDiff für ganz unterschiedliche Produkte einzusetzen, führt unweigerlich dazu, dass manche Items bei einigen Produkten als unpassend erscheinen. Ein Möglichkeit Akzeptanzprobleme zu mildern, ist das »Vorwarnen« mit einer Instruktion (siehe Kasten 1). Dabei werden die Beurteiler darauf hingewiesen, dass einige Urteile ungewohnt sein können, sie sich aber nicht so viele Gedanken darüber machen und möglichst spontan urteilen sollen. Gleichzeitig werden sie angehalten, alle Items zu beantworten. Diese Strategie kann natürlich im ungünstigen Fall dazu führen, dass Beurteiler »Verlegenheitsurteile« fällen, oder die Beantwortung ganz verweigern.

Kasten 1: Instruktion (aus 1)

»Nachfolgend finden Sie Wortpaare, mit deren Hilfe Sie das <Produkt> bewerten können. Sie stellen jeweils extreme Gegensätze dar, zwischen denen eine Abstufung möglich ist. Denken Sie nicht lange über die Wortpaare nach, sondern geben Sie bitte die Einschätzung ab, die Ihnen spontan in den Sinn kommt. Vielleicht passen einige Wortpaare nicht so gut auf das Produkt, kreuzen Sie aber trotzdem bitte immer eine Antwort an. Denken Sie daran, dass es keine »richtigen« oder »falschen« Antworten gibt – nur Ihre persönliche Meinung zählt!«

Alles in allem ist das semantische Differential als Fragebogenform recht robust. Wie bei jedem Fragebogen gilt auch hier, dass es möglichst nicht verändert werden darf. Items dürfen auf keinen Fall entfernt werden, nur weil man sie persönlich als unpassend empfindet. Es ist allerdings möglich nur einzelne Konstrukte mit Hilfe der entsprechenden Skala (I-temgruppe) zu messen.

4.0 Spezielle Eigenschaften des AttrakDiff im Vergleich

Semantische Differentiale messen die Bedeutung, die Personen einem Objekt zuweisen. Personen referenzieren dabei auf implizite Vergleichstandards, was dazu führt, dass die Wahrnehmung eines Produkts als pragmatisch oder seine Beurteilung als attraktiv zwischen den Urteilern variieren kann. Was Person A als »innovativ« (HQS) empfindet, wird von Person B vielleicht als »herkömmlich« abgetan. Damit ist das AttrakDiff explizit subjektiv, wohingegen die anderen verfügbaren Fragebögen meist objektive Produktqualität über die subjektiven Urteile der Benutzer zu erschließen suchen. Natürlich erheben so gesehen alle Fragebögen subjektive Urteile, allerdings wird Subjektivität von den meisten Fragebogenautoren als eine unerwünschte Verzerrung (Fehler) verstanden, während ich diese Varianz für bedeutsam, erklärbar und diagnostisch relevant halte.

Da das AttrakDiff im Gegensatz zu anderen Fragebögen Beurteilungen auf eher globale Produktattributen erfragt, ist es stärker erlebensorientiert. Ein Urteil wie »Die Meldungen der Software sind für mich sofort verständlich« (aus dem IsoMetrics, siehe Hamborg in diesem Band) ist sehr spezifisch. Beurteiler können es sicher erst fällen, wenn sie bereits ausreichend Erfahrungen mit dem Produkt gemacht haben. Um zu einer Antwort zu kommen, müssen Beurteiler sozusagen über eine ganze Reihe von Nutzungssituationen »mitteln«. Das koppelt die Beurteilung von kürzeren, unmittelbaren Erlebensepisoden ab. Eine globale Beurteilung wie »übersichtlich – unübersichtlich« hingehen kann sich auf einen kurzen Eindruck, eine einstündige Nutzungsepisode oder mehrere Monate der Nutzung beziehen. Kasten 2 fasst einen empirischen Hinweis auf die höhere Erlebensorientierung des AttrakDiff zusammen.

Kasten 2: Beispiel für höhere Erlebensorientierung

In einem usability test mit 15 Teilnehmern wurde nach jeder durchgeführten Aufgabe die subjektive mentale Anstrengung gemessen. Sie ist ein Indikator für erlebte Nutzungsprobleme. Pro Teilnehmer kann durch Mittelung über die Aufgaben eine mittlere Anstrengung errechnet werden. Nach dem Test wurde die Gebrauchstauglichkeit sowohl mit dem ISONORM-Fragebogen 6 als auch mit der PQ-Skala des AttrakDiff erhoben. Der ISONORM-Fragebogen erfordert wie der IsoMetrics eher spezifische Urteile. Der mittlere ISONORM-Wert korrelierte gut mit dem PQ-Wert, $Rho = .76$, was für die Gültigkeit der PQ-Skala spricht. Allerdings korrelierte der ISONORM-Wert nur niedrig mit der erlebten Anstrengung, $Rho = -.15$, während der PQ-Wert recht gut mit der Anstrengung korrelierte, $Rho = -.51$. Das tatsächlich Erlebte hatte also einen stärkeren Einfluss auf die PQ-Messung als auf die ISONORM-Messung. Sicher darf dieses Einzelergebnis nicht überbewertet werden. Es legt allerdings zumindest Unterschiede zwischen den Fragebogenformen nahe.

Das AttrakDiff ist ein, im Gegensatz zu den anderen verfügbaren Fragebogen, sehr ökonomisches Instrument. Das Ausfüllen dauert nicht länger als fünf Minuten, da eben eher eine globale, spontane als eine detaillierte Beurteilung abgegeben werden muss. Dies hat allerdings auch einen klaren Nachteil: AttrakDiff-Ergebnisse können kaum gestaltungsorientiert eingesetzt werden. Die Beurteilung eines Produkts als »schwer handhabbar« ist zwar ein eindeutiger Hinweis auf mangelnde Gebrauchstauglichkeit, den Grund für dieses Urteil kann das AttrakDiff aber nicht aufdecken. Dies steht im Gegensatz zu Fragebögen wie dem IsoMetrics (z.B. 2), der besonders in seiner formativen Version detaillierte Problemhinweise erheben kann. Ein solcher Fragebogen ist im Gegensatz zum AttrakDiff zwar weitaus aufwändiger, kann dafür aber als alleinige Maßnahme durchgeführt werden. Das AttrakDiff sollte hingegen immer mit einem Verfahren kombiniert werden, das zusätzliche gestaltungsorien-

tierte Hinweise erheben kann, wie z.B. einem usability test (siehe auch Beiträge zum usability test in diesem Band). Eine interessante Anwendung ist die Kombination des AttrakDiffs mit einem nachfolgenden laddering (siehe 7 für einen Überblick der laddering-Techniken). Dabei geht man – grob gesprochen – von einem Item des AttrakDiff aus, z.B. von »langweilig – aufregend« und verwendet die Beurteilung des Teilnehmers als Startpunkt (z.B. »langweilig«). Der Teilnehmer wird nun gebeten anzugeben, warum das Produkt als eher langweilig empfunden wird. Dabei kann zunächst nach abstrakten Kategorien gefragt werden (z.B. die Funktionen, das Aussehen, die Bedienung). In weiteren Frageschritten wird das als Startpunkt dienende Urteil auf immer konkretere Aspekte des Produkts heruntergebrochen. So können produktspezifische Ursachen für Urteile, und damit auch Ansätze für Verbesserungsmöglichkeiten, erschlossen werden.

5.0 Einbindung in Projekte

Das AttrakDiff kann im Rahmen eines Projekts recht unterschiedlich eingesetzt werden. Bei der Planung seines Einsatzes sollten die folgenden drei Aspekte geklärt werden.

Bezugsrahmen der Beurteilung.

Mit einem semantischen Differential werden im weitesten Sinne Urteile über Produkte erhoben. Menschen können normalerweise recht schnell, auch auf der Basis kurzer Eindrücke oder unvollständiger Information, Urteil bilden und auch mitteilen. Allerdings sind sie sich oft ihres Urteils weniger sicher und legen auch häufiger Wert auf die Feststellung, dass sie nicht genügend Informationen für eine »echte« Beurteilung hatten. Aus Sicht des usability professionals kann sowohl der spontane »erste Eindruck« als auch eine

auf direkte Erfahrungen mit dem Produkt basierende Beurteilung wertvoll sein. Wichtig ist es, den gewünschten Bezugsrahmen vorher zu definieren und ihn bei der Erhebung und der Interpretation zu berücksichtigen. Interessiert man sich beispielsweise für den ersten Eindruck, sollte zunächst klar sein, warum dies für das Projekt wichtig ist. Dann muss sichergestellt werden, dass sich alle Urteile auf einen vergleichbaren Eindruck beziehen, z.B. ein einminütiges, freies Explorieren des Produkts oder eine Reihe von Bildschirmausdrucken, die jeweils 30 Sekunden betrachtet werden. Bei der Interpretation muss der Bezugsrahmen immer berücksichtigt werden, denn es macht durchaus einen Unterschied, ob man aus der Anmutung der Benutzungsoberfläche Urteile über seine Gebrauchstauglichkeit ableitet oder sie in der Interaktion mit dem Produkt direkt erfährt.

Implizite oder explizite Vergleichsstandards.

Bei der Beurteilung eines Produktes mit dem AttrakDiff findet immer ein Vergleich statt. Ein Beurteiler, der ein Produkt als »innovativ« bezeichnet, vergleicht seine Eindrücke mit Standards. Diese Standards sind in der Regel implizit bzw. persönlich. Sie können aus einem oder einer Reihe von Produkten, die dem Beurteiler als Vergleichsstandard angemessen erscheinen bestehen oder aber etwas abstrakt Wünschenswertes bzw. eine Verdichtung unterschiedlicher Erfahrungen und Vorstellungen sein. Ein Softwareprodukt für den Arbeitsplatz wird beispielsweise eher spontan mit anderen Produkten aus der Arbeitswelt verglichen und nicht mit dem Lieblingscomputerspiel. Dementsprechend ist jedes positive AttrakDiff-Urteil als ein »besser, als das was der Beurteiler als Standard heranzieht« und jedes negative als ein »schlechter als der

Standard« zu interpretieren. Unter Umständen kann es sinnvoll sein, den Standard explizit zu machen. Dies kann man erreichen, indem man die Beurteiler bei der Beurteilung bittet, ein bestimmtes Produkt (z.B., eine vorherige Version des zu beurteilenden Produkts) als Vergleich heranzuziehen. Ein expliziter Standard erleichtert die Interpretation und vermindert die Unterschiedlichkeit (Varianz) der Urteile.

Produktentwürfen den »Besten«, genügt es, alle Entwürfe bewerten zu lassen und dann den mit den besten Werten zu wählen. Die Gründe für das gute Abschneiden sind hier zweitrangig. Im Rahmen des usability engineerings spielt allerdings die Verbesserung eines Produkts die zentrale Rolle. Aus diesem Grund sollte das AttrakDiff mit anderen, gestaltungsorientierten Methoden kombiniert werden. Das können sowohl usability tests, als auch andere Fragebögen oder Expertenbegutachtungen sein. Allerdings geben diese klassischen Methoden des usability engineering hauptsächlich Aufschluss über Gründe für mangelnde Gebrauchstauglichkeit. Gründe für den Mangel an hedonischen Qualitäten müssen mit anderen, u.U. noch zu entwickelnden Verfahren, erhoben werden. Einen Ansatz bietet beispielsweise das oben beschriebene laddering.

Vor der Durchführung eines Projektes mit dem AttrakDiff muss also zumindest der Bezugsrahmen der Beurteilung definiert werden.

Außerdem sollte die Wichtigkeit eines expliziten Vergleichsstandard und zusätzlich zu erhebende, qualitative Daten geklärt werden. Aus diesen Überlegungen ergeben sich unterschiedlichste Möglichkeiten der Einbindung (siehe Kasten 3 für einige Beispiele).

Kasten 3: Projekttypen

Einmalmessung des ersten Eindrucks oder des Erlebten.

Dabei wird das Produkt mit den persönlichen, impliziten Standards der Beurteiler verglichen. Werte die deutlich (d.h., statistisch signifikant) besser als 0 sind (die Skala reicht von -3 bis +3) weisen auf ein überdurchschnittliches, Werte die deutlich schlechter als 0 sind, auf ein unterdurchschnittliches Produkt hin.

Den ersten Eindruck mit dem Erlebten vergleichen.

Dabei wird das Produkt zunächst kurz vorgestellt (Tour, freie Exploration) und dann bewertet. Die zweite Bewertung findet nach dem Umgang mit dem Produkt statt. Beide Bewertungen geben Hinweise auf die Qualität des Produkts. Große Unterschiede zwischen beiden Messungen haben einen zusätzlichen diagnostischen Wert: Ist der erste Eindruck deutlich besser als die Bewertung nach Erleben, ist das Potential für Enttäuschung besonders hoch. Die Benutzungsoberfläche weckt Erwartungen, die dann nicht erfüllt werden. Ist der erste Eindruck wesentlich schlechter als die Bewertung nach Erleben, muss ein höheres Augenmerk auf die Anmutung der Benutzungsoberfläche gelegt werden. Sie kann offensichtlich die erlebte Qualität des Produkts nicht unmittelbar kommunizieren.

Vergleiche mit expliziten Standards.

Dabei wird zunächst ein Vergleichsprodukt definiert. Das kann zum Beispiel ein wichtiges Konkurrenzprodukt oder ein Vorgängerprodukt sein. Wichtig ist allerdings, dass der Beurteiler das Vergleichsprodukt kennt. Kann dies nicht vorausgesetzt werden, kann auch ein generisches Vergleichsprodukt definiert werden, wie zum Beispiel: »Das Produkt, das Sie bei der Arbeit am häufigsten nutzen«. Dieses Produkt kann zwar für jeden Beurteiler ein anderes sein, aber diese Vergleichsprodukte haben ein gemeinsames Merkmal, in diesem Fall die Nutzungshäufigkeit. Das Vergleichsprodukt wird dann zunächst beurteilt. Die Beurteilung des eigentlichen Produkts kann dann mit der Beurteilung des Vergleichsproduktes relativiert werden. Es kann durchaus sein, dass die Beurteilung des Produkts nicht eindeutig im positiven Bereich liegt, aber doch deutlich besser als das Vergleichsprodukt ist.

Mehrere Produkte miteinander verglichen.

Dies können zum Beispiel Gestaltungsalternativen des gleichen Produkts sein (welche Benutzungsoberfläche wirkt am besten?) oder Konkurrenzprodukte.

6.0 Schluss

Das AttrakDiff ist ein flexibler und ökonomischer Fragebogen zur Bewertung interaktiver Produkte. Sein Hauptvorteil liegt in der Berücksichtigung von Qualitäten, die über die Gebrauchstauglichkeit hinausgehen. Sein größter Nachteil ist die fehlende Gestaltungsorientierung.

Der Einsatz des AttrakDiff will gut geplant sein, da die Art und Weise der Einbindung in ein Projekt sowohl Auswirkungen auf die erhobenen Daten, als auch auf die Möglichkeiten und Grenzen der Interpretation hat. Bei Interesse bietet die Website www.attrakdiff.de weitere Informationen zum AttrakDiff. Gleichzeitig wird auch eine Online-Version des Fragebogens kostenlos zur Verfügung gestellt. Sie unterstützt einige der angesprochenen Projekttypen, bietet eine automatisierte statistische Auswertung und diverse Interpretationshilfen. Dies ist sicher eine gute Gelegenheit, das AttrakDiff kennen zu lernen und vielleicht in eigene Projekte einzubinden.

5.0 Literatur

- 1 J. Bortz & N. Döring: Forschungsmethoden und Evaluation für Sozialwissenschaftler. Berlin, 1995: Springer.
- 2 Gediga, G., Hamborg, K.-C. & Düntsch, I.: The IsoMetrics usability inventory: an operationalization of ISO 9241-10 supporting summative and formative evaluation of software systems. In: Behaviour & Information Technology 18 3 (1999), 151-164.
- 3 Hamborg, K.-C. & Gediga, G.: Methoden und Modelle zur Gestaltung gebrauchstauglicher Software. In: B. Zimolong und U. Konradt (Hg.): Enzyklopädie der Psychologie. Ingenieurspsychologie. Göttingen, im Druck: Hogrefe.
- 4 Hassenzahl, M.: Interaktive Produkte wahrnehmen, erleben, bewerten und gestalten. In: F. Thissen und P. F. Stephan (Hg.): Knowledge Media Design – Grundlagen und Perspektiven einer neuen Gestaltungsdisziplin. In Vorbereitung, Springer.
- 5 Hassenzahl, M., Burmester, M. & Koller, F.: AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In: J. Ziegler und G. Szwillus (Hg.): Mensch & Computer 2003. Interaktion in Bewegung. Stuttgart, 2003: Teubner. 187-196.
- 6 Prümper, J.: Der Benutzungsfragebogen ISONORM 9241/10: Ergebnisse zur Reliabilität und Validität. In: R. Liskowsky, B. M. Velichkovsky, und W. Wüschmann (Hg.): Software-Ergonomie '97. Usability Engineering: Integration von Mensch-Computer-Interaktion und Software-Entwicklung. Stuttgart, 1997: Teubner. 253-262.
- 7 Rugg, G. et al.: Eliciting information about organizational culture via laddering. In: Information Systems Journal 12 3 (2002), 215-229.

»Es ist erlaubt digitale und Kopien in Papierform des ganzen Papers oder Teilen davon für den persönlichen Gebrauch oder zur Verwendung in Lehrveranstaltungen zu erstellen. Der Verkauf oder gewerbliche Vertrieb ist untersagt. Rückfragen sind zu stellen an den Vorstand des GC-UPA e.V. (Postfach 80 06 46, 70506 Stuttgart).
 Proceedings of the 2nd annual GC-UPA Track
 Paderborn, September 2004
 © 2004 German Chapter of the UPA e.V.«





