

Erfolg von Massive Open Online Courses - messbar?

Katharina Drawert, Moritz Mager

Leuphana Universität Lüneburg

### Zusammenfassung

Dies hier ist ein Blindtext zum Testen von Textausgaben. Wer diesen Text liest, ist selbst schuld. Der Text gibt lediglich den Grauwert der Schrift an. Ist das wirklich so? Ist es gleichgültig, ob ich schreibe: „Dies ist ein Blindtext“ oder „Huardest gefburn“? Kjift – mitnichten! Ein Blindtext bietet mir wichtige Informationen. An ihm messe ich die Lesbarkeit einer Schrift, ihre Anmutung, wie harmonisch die Figuren zueinander stehen und prüfe, wie breit oder schmal sie läuft. Ein Blindtext sollte möglichst viele verschiedene Buchstaben enthalten und in der Originalsprache gesetzt sein. Er muss keinen Sinn ergeben, sollte aber lesbar sein. Fremdsprachige Texte wie „Lorem ipsum“ dienen nicht dem eigentlichen Zweck, da sie eine falsche Anmutung vermitteln.

1. Function: An abstract summarizes, in one paragraph (usually), the major aspects of the entire paper in the following prescribed sequence:

the question(s) you investigated (or purpose), (from Introduction) state the purpose very clearly in the first or second sentence. the experimental design and methods used, (from Methods) clearly express the basic design of the study. Name or briefly describe the basic methodology used without going into excessive detail-be sure to indicate the key techniques used. the major findings including key quantitative results, or trends (from Results) report those results which answer the questions you were asking identify trends, relative change or differences, etc. a brief summary of your interpetations and conclusions. (from Discussion) clearly state the implications of the answers your results gave you.

*Keywords:* MOOC, SEM, IS Success

## Einleitung

Dies hier ist ein Blindtext zum Testen von Textausgaben. Wer diesen Text liest, ist selbst schuld. Der Text gibt lediglich den Grauwert der Schrift an. Ist das wirklich so? Ist es gleichgültig, ob ich schreibe: „Dies ist ein Blindtext“ oder „Huardest gefburn“? Kjift – mitnichten! Ein Blindtext bietet mir wichtige Informationen. An ihm messe ich die Lesbarkeit einer Schrift, ihre Anmutung, wie harmonisch die Figuren zueinander stehen und prüfe, wie breit oder schmal sie läuft. Ein Blindtext sollte möglichst viele verschiedene Buchstaben enthalten und in der Originalsprache gesetzt sein. Er muss keinen Sinn ergeben, sollte aber lesbar sein. Fremdsprachige Texte wie „Lorem ipsum“ dienen nicht dem eigentlichen Zweck, da sie eine falsche Anmutung vermitteln.

## Grundlagen

Es war das Jahr 2008, als George Siemens und Stephen Downes an der Universität Manitoba in Kanada eine Vorlesung über das Internet verbreiteten und dabei über 2200 Teilnehmer von der ganzen Welt erreichen konnten. Dieses Ereignis ging als der erste Massive Open Online Course (im folgenden MOOC genannt) in die Geschichte ein. Jeder Interessierte konnte damals ohne jegliche Kosten an dem Kurs teilnehmen. Diese Idee entwickelte sich in den folgenden Jahren weiter und ab 2011 begannen die ersten US-amerikanischen Universitäten den eigenen Studierenden MOOCs als Erweiterungsangebot anzubieten. Seit dem bieten die meisten Universitäten eigene MOOCs an, welche sich dabei meist an externe Personen richten. Darüber hinaus gibt es unterschiedliche Unternehmen, die eigene MOOCs anbieten. In der Regel ist die Teilnahme an einem MOOC kostenlos, die Ausstellung eines Zertifikats inklusive Credit Points nach erfolgreicher Teilnahme jedoch nur gegen Gebühr möglich.

Auch die Leuphana Universität Lüneburg bietet mit der Digital School Massive Open Online Courses an. Dabei werden unterschiedliche Kooperationspartner wie unter anderem das Goethe-Institut mit einbezogen.

## Modell

In this section, the research variables and hypotheses are presented.

Table 2 Definitions of dimensions.

Tabelle 1

### *Definition der Dimensionen*

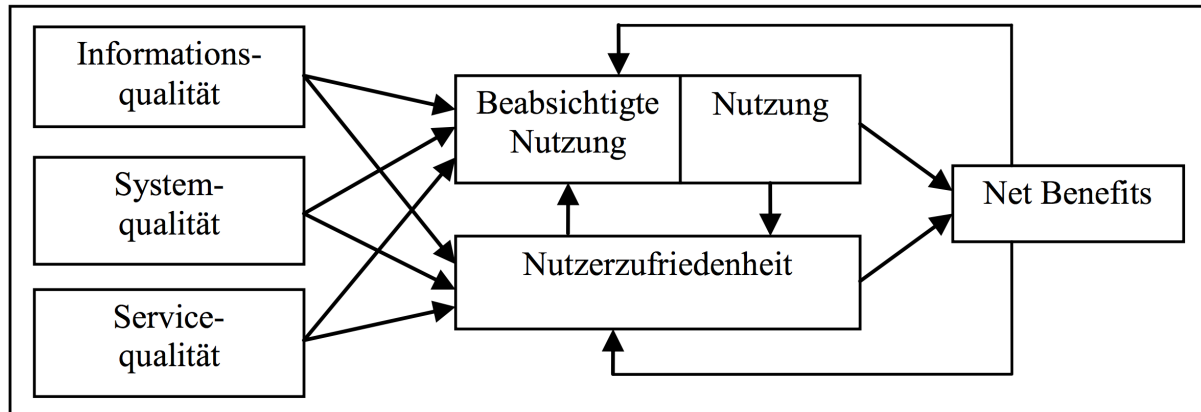
Konstrukt	Definiton	Quelle
Servicequalität	Qualitätsfaktor für die erwarteten Support,den die Nutzer in Anspruch nehmen können	<b>petter2008measuring</b>
Systemqualität	Die erwarteten Eigenschaften und Funktionen von dem System	<b>petter2008measuring</b>
Nutzerzufriedenheit	Das Ausmaß darüber, in wie weit die Bedürfnisse, Ziele und Wünsche während des MOOC erfüllt werden	<b>sanchez2009moderating</b>
Net Benefit	Drück aus, in wie weit Informationssysteme zum Erfolg einzelner Personen, Gruppen und Organisation beitragen können	<b>conf/gi/GemlikNSB10</b> <b>Petter:0aa</b>

## Empirische Datenbasis

### Forschungsmethode

Für die Untersuchung der Hypothesen und der Darstellung der Ergebnisse wurde während eines Massive Open Online Course zu drei Zeitpunkten ein Fragebogen an die Teilnehmer versendet. Schon bei der Konzeption des Fragebogens musste das Modell für die anschließende Auswertung der Daten feststehen, da die Fragen entsprechend gestellt werden mussten. In diesem Kontext wird dazu häufig ein Modell von DeLone und McLean angewandt, welches international unter dem Namen „IS Success Model“ bekannt ist. Die Erfolgsmessung von Informationssystemen verfolgt dabei einem bestimmten Muster und ermöglicht somit Vergleiche mit anderen Erhebungen. Seit der ersten Entwicklung im Jahr 1992 wurde das Modell intensiv diskutiert und dabei empirisch auf die Qualität hin überprüft. Grundlegend stellten DeLone und McLean fest, dass sich fast alle Erfolgsmessungen in nur sechs Kategorien einordnen lassen, die untereinander als abhängige

Variablen dargestellt werden können. Im Laufe der wissenschaftlichen Weiterentwicklung wird in der Literatur das aktuelle Modell wie in Abbildung grafisch dargestellt. (vgl. **delone2002information**)



Das im Rahmen dieses Papers verwendete IS Success Modell beinhaltet allerdings nur 4 Konstrukte, da erstens die Anzahl der Fragen streng reglementiert war und darüber hinaus auch nur eine geringe Response Rate erwartet wurde. Somit ist es mit dem kleinen sample besser möglich, qualitativ gute Aussagen zu treffen. Die Konstrukte sind in Tabelle XX 3 genau definiert. Aufgrund der internationalen Ausrichtung des Kurses wurden die Fragen in Englisch gestellt, wobei eine Seven-point Likert Skala verwendet wurde. Die Antwortmöglichkeiten reichten von „Strongly disagree (1)“ bis „Strongly Agree (7)“. Wie eingehend schon erwähnt, fand die Erhebung der Daten zu drei unterschiedlichen Zeitpunkten statt: zu Beginn des Kurses (T1), während des Kurses (T2) und zum Ende des Kurses (T3). Die Beantwortung der Umfrage unterlag einer freiwilligen Basis. Die im Forschungsmodell beschriebenen Items sind in den Fragebögen T2 und T3 enthalten. Für die Erstellung des Modells konnten jedoch nur Daten aus T2 und T3 verwendet werden, da die Teilnehmer zu Beginn des Kurses keine Angaben über ihren persönlichen Erfolg machen konnten (Frageitem: Enriched Knowledge fehlt in T1).

ITEM Enri  
knowledge  
nau benenn

Teilnehmer Die vorliegenden Daten entstammen aus Befragungen der Teilnehmer des MOOCs (Mentored Open Online Course) Psychology of Negotiations - Reaching Sustainable Agreements in Negotiations on "Commons" der Leuphana Digital School. Der Kurs fand von Mai bis August 2014 statt und war offen für Teilnehmer aus der ganzen

Tabelle 2

*Forschungsmodell*

Konstrukt	Item	Factorloadings
Servicequalität	The Leuphana Digital School provides a proper level of online assistance and explanation	0,794
	The teaching staff is highly availability for consultation	0,913
	The teaching staff provides satisfactory support to users using Leuphana Digital School	0,887
Systemqualität	Leuphana Digital School's technical system has attractive features to appeal to the users.	0,895
	Leuphana Digital School's technical system is easy to use.	0,865
	Leuphana Digital School's technical system provides a personalized information presentation.	0,808
Nutzerzufriedenheit	Most of the users bring a positive attitude or evaluation towards Leuphana Digital School.	0,763
	Leuphana Digital School's technical system is easy to use.	0,865
Net Benefit	Leuphana Digital School helps you think through problems.	0,927
	All in all, my knowledge has been enriched as a result of the course	0,768

Welt. Die Teilnehmerzahl wurde auf 1000 begrenzt um die Qualität der Beratung und Führung durch Mentoren und Lehrende zu gewährleisten. Der Kurs war gebührenfrei. Bei erfolgreichem Abschluss konnten Teilnehmer - gegen eine Gebühr von 20 €- ein Zertifikat der Universität erhalten und 5 Credit Points (ECTS) zugeschrieben bekommen, welche dem eigenen Studium angerechnet werden konnten. Die Fragebögen wurden an alle Teilnehmer verschickt. Da die Anzahl der Kursteilnehmer, die den Fragebogen erhalten haben unbekannt ist, kann die Returnquote nicht angegeben werden. Gemessen an den Antworten, ist diese jedoch eher gering - vor allem in T2 und T3. Bei Ersterem lagen 32 Antworten vor, wovon nach Bereinigung von ungültigen oder unvollständigen Antworten 29 verwertbare waren. Die Bereinigung ungültiger und unvollständiger Antworten reduzierte die nutzbaren Antworten in T3 von 48 auf 36. Insgesamt liegen damit 65 Datensätze für die Modellüberprüfung vor, was einen relativ kleinen Stichprobenumfang

darstellt. **jakobowicz2006understanding** gaben an, dass für ein komplexes Modell ein Stichprobenumfang von mindestens 200 empfehlenswert ist, da ein geringer Stichprobenumfang zu Verzerrungen bei der Parameterschätzung und einem großem Standardfehler führen. Das in dieser Arbeit überprüfte Modell ist mit zwei unabhängig latenten (Systemqualität und Servicequalität) und zwei abhängig latenten Variablen (Nutzerzufriedenheit und Persönlicher Nutzen) verhältnismäßig einfach, daher kann auch ein kleiner Stichprobenumfang ausreichend Erkenntnisse liefern, gleichzeitig sollte bei der Interpretation der Ergebnisse jedoch der Stichprobenumfang berücksichtigt werden. Die Demographischen Daten der Fragebögen T2 und T3 lassen sich aus Tabelle 4 entnehmen.

**Data Analysis Technique Instrument validation** We followed the procedures outlined by Gefen and Straub (2005) to test discriminant and convergent validity. Discriminant validity refers to whether the items measure the construct in question or other (related) constructs (Gefen and Straub, 2005). We verified discriminant validity using correlation matrix and factor analysis. Table 4 shows the correlation matrix with the square root of average variance extracted (AVE) values presented diagonally. The square root of the AVE value for the variables is consistently greater than the off-diagonal correlation values, suggesting satisfactory discriminant validity between the variables (Fornell and Larcker, 1981)

Zur Analyse im Rahmen dieser Arbeit wurde auf Strukturgleichungsmodellierung zurückgegriffen, For this paper, structural equation modeling (SEM) using partial least squares (PLS) was used to evaluate the research model and hypotheses.

## Ergebnisse

Zur Evaluierung des Forschungsmodells und der Hypothesen wurde die Partial-Least-Square(PLS)-Methode verwendet. Dabei werden „die Modellparameter so geschätzt, dass der Anteil der erklärten Varianz der abhängigen Variable und der Indikatoren eines reflektiv gemessenen Konstrukts maximiert wird“ (**nitzl2010anwenderorientierte**). Ein besonderer Vorteil der PLS-Methode ist ihre Anwendungsmöglichkeit auch bei verhältnismäßig kleiner Stichprobengröße. Zur Kalkulation einer minimalen Stichprobengröße

kommt häufig eine Faustregel zur Anwendung, nach der die Stichprobengröße mindestens das zehnfache des Konstruktes mit der größten Anzahl zu schätzender Parameter sein sollte (**islam2013investigating**). Dieses Kriterium wird in dieser Studie erfüllt. Zur Analyse wurde die Softwareapplikation SmartPLS verwendet.

Für die Modellbeurteilung wird zunächst das reflektive Messmodell einer Güteprüfung unterzogen. Die Konvergenzvalidität wird anhand der Kriterien Indikatorreliabilität, Konstruktreliabilität und durchschnittlich erfassten Varianz (DEV) kritisch betrachtet, während die Validität mithilfe der Diskriminanzvalidität überprüft wird.

Die Indikatorreliabilität testet, ob sich ein Indikator für die Messung einer latenten Variable eignet. Eine Faktorladung  $\lambda > 0,7$  gilt als signifikant (**nitzl2010anwenderorientierte**). Der Wert wird von allen Indikatoren erreicht (siehe Tabelle 3).

Die Konstruktreliabilität  $\rho$  untersucht unter Einsatz der internen Konsistenz (auch composite reliability (CR) genannt) „wie gut die Indikatoren eine latente Variable wiedergeben“ (**nitzl2010anwenderorientierte**). Ein Wert von  $\rho \geq 0,6$  gilt als akzeptabel (**ringle2007beurteilung**). Das Messmodell weist  $\rho > 0,8$  auf und liegt damit über den Schwellenwert.

Die durchschnittliche erfasste Varianz (DEV) „setzt den Anteil der erklärten Varianz in Relation zum Messfehler einer latenten Variable“ (**nitzl2010anwenderorientierte**). Ein Wert von  $DEV \geq 0,5$  stellt einen ausreichend hohen Wert dar. Die DEV liegt in dieser Studie mit  $DEV > 0,7$  ebenfalls über dem genannten Schwellenwert und ist somit akzeptabel. Die entsprechenden Werte für die Konstruktreliabilität und die DEV können Tabelle 5 entnommen werden.

Die Diskriminanzvalidität hingegen „gibt an, in welchem Ausmaß sich die Indikatoren eines Konstrukts von denen eines anderen Konstrukts unterscheiden“ (**nitzl2010anwenderorientierte**). Zur Überprüfung der Diskriminanzvalidität kann das Fornell-Larcker-Kriterium und die Cross Loadings herangezogen werden. Bei ersterem wird die Wurzel der DEV einer latenten Variable verglichen mit jeder Korrelation dieser latenten Variable mit einer anderen latenten Variablen und sollte stets größer sein (**nitzl2010anwenderorientierte**). Das Fornell-Larcker-Kriterium wird in dieser Studie erfüllt (siehe Tabelle 6). Die Cross Loa-



dings können Tabelle 7 entnommen werden. Ein Indikator sollte dabei die stärkste Beziehung mit dem ihm zugeordneten Konstrukt aufweisen (**nitzl2010anwenderorientierte**), was ebenfalls erfüllt ist.

Das Messmodell erfüllt damit alle Gütekriterien. Zur Beurteilung des Strukturmodells werden das Bestimmtheitsmaß  $R^2$ , die Pfadkoeffizienten, die Effektstärke  $f^2$  und die Prognoserelevanz  $Q^2$  herangezogen.

Die Werte für das Bestimmtheitsmaß  $R^2$  sind in Tabelle 5 enthalten. Das Bestimmtheitsmaß „gibt den Anteil der erklärten Varianz im Verhältnis zur Gesamtvarianz an.“ (**nitzl2010anwenderorientierte**) Eine Einteilung relevanterer Schwellenwerte wurde von **chin1998partial** in einer Studie ermittelt. Die Werte für  $R^2$  von 0,67, 0,33 und 0,19 wurden in „substanziell“, „mittelgut“ und „schwach“ eingeteilt. In dieser Studie sind die  $R^2$  dementsprechend als „mittelgut“ (Nutzerzufriedenheit: 0,387 / persönlicher Erfolg: 0,369) einzustufen.

Als Schwellenwerte für die Effektstärke  $f^2$  wurde von **chin1998partial** 0,02, 0,15 bzw. 0,35 ermittelt, die aussagen, ob eine unabhängige latente Variable einen geringen, mittleren bzw. großen Einfluss auf eine abhängige latente Variable hat. Demnach weist Nutzerzufriedenheit auf den persönlichen Lernerfolg ( $f^2=0,306$ ) einen mittleren Einfluss und Systemqualität auf Nutzerzufriedenheit ( $f^2=0,519$ ) einen großen Einfluss aus.

Die Pfadkoeffizienten  $\gamma$  geben die Stärke der Kausalbeziehung zwischen den latenten Variablen an. Sie können Werte zwischen -1 und 1 annehmen. Ein Wert Nahe 0 gilt als schwach. Als signifikant wird ein Wert kleiner -0,2 oder größer 0,2 angesehen (**chin1998commentary**). In dieser Studie haben dementsprechend die Nutzerzufriedenheit auf den persönlichen Erfolg ( $\gamma = 0,557$ ;  $p < 0,001$ ) und die Servicequalität auf die Nutzerzufriedenheit ( $\gamma = 0,591$ ;  $p < 0,001$ ) einen signifikant positiven Einfluss. (siehe Abbildung...) Die mithilfe der Bootstrapping-Methode ermittelten T-Werte bestätigen die Signifikanz der beiden Pfadkoeffizienten. Die restlichen Hypothesen können hingegen nicht bestätigt werden.

Das Geisser-Stone-Kriterium sieht eine ausreichende Prognoserelevanz wenn  $Q^2 > 0$ , was in dieser Studie erfüllt wurde (siehe Tabelle 5)

Die Prognoserelevanz  $Q^2$  wird nach dem Geisser-Stone-Kriterium überprüft, wonach man bei  $Q^2 > 0$  von einer ausreichenden Prognoserelevanz spricht. Die Berechnung erfolgt mithilfe der Blindfolding-Methode in SmartPLS, die Ergebnisse lassen sich Tabelle 5 entnehmen

## Interpretation

Sed feugiat. Cum socii natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Ut pellentesque augue sed urna. Vestibulum diam eros, fringilla et, consectetuer eu, nonummy id, sapien. Nullam at lectus. In sagittis ultrices mauris. Curabitur malesuada erat sit amet massa. Fusce blandit. Aliquam erat volutpat. Aliquam euismod. Aenean vel lectus. Nunc imperdiet justo nec dolor.

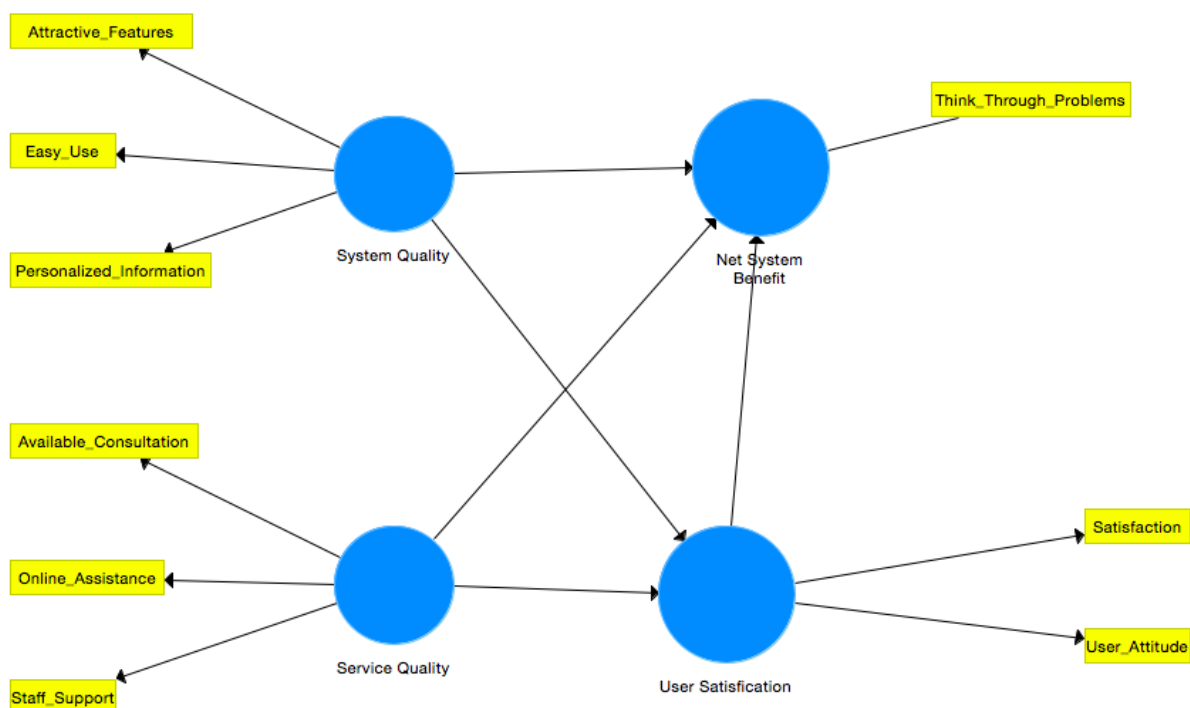


Abbildung 1. Research Model

## Abbildungsverzeichnis

1	Research Model . . . . .	10
---	--------------------------	----

**Tabellenverzeichnis**

1	Definition der Dimensionen . . . . .	4
2	Forschungsmodell . . . . .	6
3	Forschungsmodell2 . . . . .	12
4	Demographische Daten . . . . .	13
5	Übersicht Gütekriterien . . . . .	13
6	Fornell-Larcker-Kriterium . . . . .	14
7	Cross Loadings . . . . .	14

Tabelle 3

*Forschungsmodell2*

<b>Konstrukt</b>	<b>Item</b>
Servicequalität	The Leuphana Digital School provides a proper level of online assistance and explanation
	The teaching staff is highly availability for consultation
	The teaching staff provides satisfactory support to users using Leuphana Digital School
Systemqualität	Leuphana Digital School's technical system has attractive features to appeal to the users.
	Leuphana Digital School's technical system is easy to use.
	Leuphana Digital School's technical system provides a personalized information presentation.
Nutzerzufriedenheit	Most of the users bring a positive attitude or evaluation towards Leuphana Digital School.
	Leuphana Digital School's technical system is easy to use.
Persönlicher Nutzen	Leuphana Digital School helps you think through problems.
	All in all, my knowledge has been enriched as a result of the course (nur in Fragebogen 2 und 3)

Tabelle 4

*Demographische Daten*

	Anzahl	in Prozent
<i>Geschlecht</i>		
weiblich	40	62
männlich	25	38
Total	65	100
<i>Alter</i>		
21-30 Jahre	35	54
31-40 Jahre	10	15
> 41 Jahre	20	31
Total	65	100
<i>Herkunft</i>		
Deutschland	27	42
Europa (excl. Deutschland)	18	28
Afrika	6	9
Asien	6	9
Nordamerika	4	6
Südamerika	4	6
Total	65	100

Tabelle 5

*Übersicht Gütekriterien*

Faktor	DVE	CR	R <sup>2</sup>	Q <sup>2</sup>	f <sup>2</sup>
Net Benefit	0,725	0,839	0.503	0.476	1.000
Servicequalität	0,750	0,900			0.882
Systemqualität	0,734	0,892			0.817
Nutzerzufriedenheit	0,703	0,824	0.526	0.388	

Tabelle 6

*Fornell-Larcker-Kriterium*

	Net Benefit	Servicequalität	Systemqualität	Nutzerzufriedenheit
Net Benefit	0,725	0,839	0.503	0.476
Servicequalität	0,750	0,900		
Systemqualität	0,734	0,892		
Nutzerzufriedenheit	0,703	0,824	0.526	0.388

Tabelle 7

*Cross Loadings*

Faktor	DVE	CR	R <sup>2</sup>	Q <sup>2</sup>	f <sup>2</sup>
Net Benefit	0,725	0,839	0.503	0.476	1.000
Servicequalität	0,750	0,900			0.882
Systemqualität	0,734	0,892			0.817
Nutzerzufriedenheit	0,703	0,824	0.526	0.388	