Universität Konstanz

Fachbereich für Politik- und Verwaltungswissenschaft

Diplomarbeit

Evaluation in öffentlichen Dienstleistungsorganisationen

Eine empirische Untersuchung zur Entstehung des Publikumsurteils im Musiktheater

Erstgutachterin: Prof. Dr. Sabine Boerner

Zweitgutachter: Prof. Dr. Philip Manow

Johanna Jobst

Allensteinerstraße 9

78467 Konstanz

Matrikel-Nr.: 01/529391

Johanna.Jobst@uni-konstanz.de

Konstanz, den 18. September 2007

Konstanzer Online-Publikations-System (KOPS)
URL: http://www.ub.uni-konstanz.de/kops/volltexte/2007/3846/
URN: http://nbn-resolving.de/urn:nbn:de:bsz:352-opus-38468

Abstract

Within the framework of the comprehensive reforms known as ‚New Public Management', the issue of measuring performance quality in public opera companies comes to the fore. Since quality in opera companies is generally considered difficult to measure (Heinze, 1994; Lipp, 1994; Hoegl, 1995; Allmann, 1997) and quality concepts both for for-profit and nonprofit companies are not applicable (Boerner, 2002), there exists no valid and reliable scientific tool that is able to grasp the performance quality of an opera company. Therefore the aim of this paper is to close this gap by conducting research in the evaluation of opera performances.

By using a modified version of the questionnaire that is based on Boerner's componential framework of performance quality in the opera and was developed by Moser (2005) und Renz (2006) to understand the perception of performance quality in the opera this study examines two objects: First of all it investigates the content, the structure and the agreement in judgments on performance quality in the opera. The second part of this work focuses on the audience's judgment on the whole evening in the opera as the analysis of the literature suggests that there are other criteria besides the performance quality used by the audience. Members of the audience of a performance of "Die Zauberflöte" by Wolfgang Amadeus Mozart were asked to answer the questionnaire after attending a performance in the opera house of Dessau (n = 120). In terms of the performance's content, factor analysis yields that the audience separates between the fit between the musical dimension and the staging dimension, the fit within the musical dimension, the fit within the staging dimension, the orchestra, the choir, the soloists and the staging/scenery. A following path analysis confirms a hierarchical structure of the perception of performance quality as assumed by the componential framework of Boerner (2002). Furthermore, the fit between the musical dimension and the staging dimension is shown to be the most important component with regard to the weight within the audience's judgment concerning the performance quality. The agreement about the assessment within both the sample of non-experts and of experts is high. While identifying the determinants of the audience's judgment on the whole evening in the opera house, path analysis yields the affective response to the performance as a better predictor of enjoyment than the perceived quality of the performance. Additional variables (such as the complexity of the production or the audience's satisfaction with the auditorium) were shown to make virtually no contribution to the audience's overall enjoyment of the evening in the opera. Again, the agreement within the sample of both the non-experts and the experts about the assessment of the different aspects is generally high.

Inhaltsverzeichnis

1. EINLEITUNG 1

1.1 Relevanz und Ziele der Arbeit 1

1.2 Vorgehen 4

2. FRAGESTELLUNG DER ARBEIT, STAND DER FORSCHUNG UND EIGENE ÜBERLEGUNGEN 5

2.1 Untersuchungsfragen 5

2.2 Stand der Forschung 7
2.2.1 Das analytische Modell zur künstlerischen Qualität im Musiktheater 7
2.2.2 Empirische Überprüfung des analytischen Modells 11
2.2.3 Erklärungen individueller Reaktionen im Musiktheater 13
2.2.3.1 Erklärungen individueller Urteile über einen Abend im Konzert 13
2.2.3.1.1 Das Gesamterlebnis ‚Konzertaufführung' 14
2.2.3.1.2 Der Stimulus ‚aufgeführtes Werk' 16
2.2.3.1.3 Der Zuhörer 18
2.2.3.1.4 Die Situation 20
2.2.3.2 Erklärungen individueller Urteile über einen Abend im Sprechtheater 20
2.2.3.2.1 Der Stimulus ‚Aufführung im Sprechtheater' 21
2.2.3.2.2 Die emotionale Wirkung einer Theateraufführung auf den Zuschauer 22
2.2.3.2.3 Die Situation 25

2.3 Determinanten des Gesamturteils des Publikums über einen Abend in der Oper (‚Gesamterlebnis Opernabend') 26
2.3.1 Die wahrgenommene künstlerische Qualität 26
2.3.2 Individuelle Reaktion des Zuschauers 27
2.3.3 Situative Rahmenbedingungen 28
2.3.4 Das ‚Gesamterlebnis Opernabend' 28

3. INSTRUMENT UND METHODE 30

3.1 Entwicklung des Fragebogens zur Erfassung der Determinanten des Gesamturteils des Publikums über einen Abend in der Oper (‚Gesamterlebnis Opernaufführung') 30
3.1.1 Aufbau und Antwortformat des Fragebogens 30
3.1.2 Modifikation der Skalen und Items 34

3.2 Datenerhebung 38

3.3 Vorgehensweise der Datenauswertung 40

4. VORBEREITENDE ANALYSEN 42

4.1 Deskriptive Statistiken 42
4.1.1 Beschreibung der Stichprobe 42
4.1.2 Varianzanalysen 43
4.1.3 Deskriptive Auswertung 45

4.1.4 Bildung des Indexes ‚Expertentum' 54

4.2 Reliabilität und Validität des Fragebogens 55
4.2.1 Reliabilitätsanalysen 56
4.2.2 Interkorrelation der Skalen 59
4.2.3 Faktorenanalysen 63

5. ANALYSEN ZUR KLÄRUNG DER UNTERSUCHUNGSFRAGEN 71

5.1 Ergebnisse zu Untersuchungsfrage 1: Gegenstände der Bewertung künstlerischer Qualität in der Oper 71
5.1.1 Untersuchungsfrage 1a: musikalische und szenische Aspekte einer Opernaufführung 71
5.1.2 Untersuchungsfrage 1b: Gewichtung der Teilaspekte 72

5.2 Ergebnisse zu Untersuchungsfrage 2: Unterschiede zwischen Laien und Experten in der Bewertung der künstlerischen Qualität einer Opernaufführung 85
5.2.1 Untersuchungsfrage 2a: Unterschiede zwischen Laien und Experten hinsichtlich Differenziertheit und Gewichtung 86
5.2.2 Untersuchungsfrage 2b: Unterschiede zwischen Laien und Experten in der Bewertung der Teilaspekte 90
5.2.3 Untersuchungsfrage 2c: Übereinstimmung der Urteile jeweils innerhalb der Beurteilergruppen 92

5.3 Ergebnisse zu Untersuchungsfrage 3: Determinanten des Gesamturteils des Publikums über einen Abend in der Oper (‚Gesamterlebnis Opernabend') 96

5.4 Ergebnisse zu Untersuchungsfrage 4: Unterschiede zwischen Laien und Experten in der Bewertung des ‚Gesamterlebnisses Opernabend' 98
5.4.1 Untersuchungsfrage 4a: Unterschiede zwischen Laien und Experten in der Bewertung der Teilaspekte 98
5.4.2 Untersuchungsfrage 4b: Übereinstimmung der Gesamturteile innerhalb der beiden Beurteilergruppen 100
5.4.3 Untersuchungsfrage 4c: Unterschiede zwischen Laien und Experten in der Gewichtung der Teilaspekte 101

6. SCHLUSSBETRACHTUNG 104

6.1 Zusammenfassung der Ergebnisse 104

6.2 Diskussion der Ergebnisse 108

Anhang

Anhang A: Fragebogen zur Bewertung von Aufführungen in der Oper ..... a

Anhang B: Fragebogen zur Bewertung von Aufführungen im Musiktheater – Renz (2006) ..... n

Anhang C: Einfaktorielle Varianzanalysen ..... dd

Anhang D: Reliabilitäten der theoretischen Konstrukte im Vergleich ..... ff

Anhang E: Messmodell der Teilstruktur II (Stimmigkeit innerhalb der szenischen Dimension) ..... hh

Anhang F: Messmodell der Teilstruktur III (Stimmigkeit innerhalb der musikalischen Dimension) ii

Anhang G: Wichtigkeit verschiedener Akteure für die künstlerische Qualität einer Opernaufführung im Allgemeinen ..... jj

Anhang H: Faktorenanalysen zu Untersuchungsfrage 3 ..... kk

Anhang I: Informationsmaterial für die Teilnehmer der Befragung ..... mm

Anhang J: Häufigkeiten der Antwortkategorien „habe ich nicht beachtet“ und „kann ich nicht beurteilen“ getrennt nach Laien und Experten ..... nn

Anhang K: Prozentuale Häufigkeit nicht beantworteter Items getrennt nach Laien und Experten ..... pp

Anhang L: Interkorrelation der ausgewählter Skalen ..... ss

Anhang M: Faktorenanalysen getrennt nach Laien und Experten ..... tt

Literaturverzeichnis

Literaturverzeichnis xx

Abbildungsverzeichnis

Abbildung 1: Hierarchisches Modell des Qualitätsurteils im Musiktheater 10

Abbildung 2: Modell der Determinanten des Gesamterlebnisses einer Aufführung 15

Abbildung 3: Determinanten des Gesamturteils des Publikums über einen Abend in der Oper (‚Gesamterlebnis Opernabend') 29

Abbildung 4: Aufbau des Fragebogens 31

Abbildung 5: Hierarchisches Pfadmodell des Qualitätsurteils in der Oper 75

Abbildung 6: Generalfaktormodell des Qualitätsurteil in der Oper 75

Abbildung 7: Teilstruktur I des Strukturgleichungsmodells des Urteils über die künstlerische Qualität 79

Abbildung 8: Modifiziertes Pfadmodell des Qualitätsurteils in der Oper 83

Abbildung 9: Hierarchisches Pfadmodell des Qualitätsurteils in der Oper – Laien 88

Abbildung 10: Hierarchisches Pfadmodell des Qualitätsurteils in der Oper – Experten 88

Abbildung 11: Determinanten des Gesamturteils des Publikums über einen Abend in der Oper (‚Gesamterlebnis Opernabend') 97

Abbildung 12: Determinanten des Gesamturteils des Publikums über einen Abend in der Oper (‚Gesamterlebnis Opernabend') – Laien 102

Abbildung 13: Determinanten des Gesamturteils des Publikums über einen Abend in der Oper (‚Gesamterlebnis Opernabend') – Experten 102

Abbildung 14: Teilstruktur II des Strukturgleichungsmodells des Urteils über die künstlerische Qualität ..... hh

Abbildung 15: Teilstruktur III des Strukturgleichungsmodells des Urteils über die künstlerische Qualität ii

Tabellenverzeichnis

Tabelle 1: Kombination von Potential- und Ergebnisfaktoren der musikalischen und der szenischen Dimension der Interpretationsqualität im Musiktheater 8

Tabelle 2: Hierarchisches fit-Modell der Interpretationsqualität im Musiktheater 10

Tabelle 3: Häufigkeitsverteilung der Ausfüll-Zeitpunkte 43

Tabelle 4: Einfaktorielle Varianzanalysen (Faktorvariable: Zeitpunkt des Ausfüllens) – ausgewählte Skalen 45

Tabelle 5: Prozentuale Häufigkeit der nicht beantworteten Items – ausgewählte Konstrukte 47

Tabelle 6: Wichtigkeit der verschiedenen Akteure für die künstlerische Qualität einer Opernaufführung 50

Tabelle 7: Prozentuale Häufigkeit der nicht beantworteten Items – ausgewählte Konstrukte (Laien – Experten) 51

Tabelle 8: Häufigkeitsverteilung der Punkte zur Bildung des Indexes ‚Expertentum' 55

Tabelle 9: Reliabilität der gebildeten Skalen 57

Tabelle 10: Interkorrelation der Skalen 60

Tabelle 11: Faktorenanalyse der ersten Ebene der Oper „Die Zauberflöte" 64

Tabelle 12: Faktorenanalyse der zweiten Ebene der Oper „Die Zauberflöte" 65

Tabelle 13: Faktorenanalyse der zweiten Ebene der Oper „Aida" – Renz (2006) 66

Tabelle 14: Faktorenanalyse der dritten Ebene der Oper „Die Zauberflöte" 67

Tabelle 15: Skalenbildung der Faktoren 72

Tabelle 16: Korrelationsmatrix der Skalen 73

Tabelle 17: Anpassungsgüte des Modellvergleichs 76

Tabelle 18: Anpassungsgüte der Teilstruktur I des Gesamtmodells über das Urteil über die künstlerische Qualität einer Opernaufführung 80

Tabelle 19: Anpassungsgüte der Teilstruktur II des Gesamtmodells für das Urteil über die künstlerische Qualität einer Opernaufführung 81

Tabelle 20: Anpassungsgüte des Modellvergleichs 84

Tabelle 21: Standardisierte totale Effekte des modifizierten Pfadmodells 85

Tabelle 22: Standardisierte totale Effekte des hierarchischen Pfadmodells - Laien 89

Tabelle 23: Standardisierte totale Effekte des hierarchischen Pfadmodells - Experten 89

Tabelle 24: Vergleich der Qualitätsunterschiede von Experten und Laien 91

Tabelle 25: Berechnung der Intraklassenkorrelationen 93

Tabelle 26: Vergleich der Bewertung des Gesamterlebnisses von Laien und Experten 99

Tabelle 27: Berechnung der Intraklassenkorrelationen 100

Tabelle 28: Einfaktorielle Varianzanalysen (Faktorvariable: Zeitpunkt des Ausfüllens) ..... dd

Tabelle 29: Reliabilitäten der theoretischen Konstrukte der Studien von Renz (2006), Moser (2005) und Boerner et al. (in Druck) ..... ff

Tabelle 30: Wichtigkeit der verschiedenen Akteure für die künstlerische Qualität einer Opernaufführung – Laien ..... jj

Tabelle 31: Wichtigkeit der verschiedenen Akteure für die künstlerische Qualität einer Opernaufführung – Experten ..... jj

Tabelle 32: Faktorenanalyse der abhängigen Variablen ‚Gesamturteil' ..... kk

Tabelle 33: Faktorenanalyse der unabhängigen Variablen ‚künstlerische Qualität' ..... kk

Tabelle 34: Faktorenanalyse der unabhängigen Variablen ‚emotionale Wirkung' ..... kk

Tabelle 35: Faktorenanalyse der unabhängigen Variablen ‚Komplexität' ll

Tabelle 36: Faktorenanalyse der unabhängigen Variablen ‚Vertrautheit' ll

Tabelle 37: Faktorenanalyse der unabhängigen Variablen ‚Identifikation' ll

Tabelle 38: Faktorenanalyse der unabhängigen Variablen ‚situative Rahmenbedingungen' ll

Tabelle 39: Häufigkeit der Antwortkategorien „habe ich nicht beachtet" und „kann ich nicht beurteilen" getrennt nach Laien und Experten ..... nn

Tabelle 40: Prozentuale Häufigkeit nicht beantworteter Items getrennt nach Laien und Experten ..... pp

Tabelle 41: Interkorrelation der Skalen ‚musikalische Leistung des Orchesters', ‚Zusammenspiel innerhalb des Orchesters' und ‚Leistung des Dirigenten' ..... ss

Tabelle 42: Interkorrelation der Skalen ‚Stimmigkeit innerhalb der szenischen Dimension' und ‚Stimmigkeit der Ausstattung' ..... ss

Tabelle 43: Faktorenanalyse der ersten Ebene - Laien ..... tt

Tabelle 44: Faktorenanalyse der ersten Ebene - Experten ..... tt

Tabelle 45: Faktorenanalyse der zweiten Ebene – Laien ..... tt

Tabelle 46: Faktorenanalyse der zweiten Ebene – Experten ..... uu

Tabelle 47: Faktorenanalyse der dritten Ebene - Laien ..... uu

Tabelle 48: Faktorenanalyse der dritten Ebene – Experten vv

1. Einleitung

> *Die Welt des Theaters besteht aus einer beleuchteten und einer verdunkelten Hälfte. Ist die beleuchtete Hemisphäre der Theaterwelt, auf der so viele Faktoren und Aktivitäten zusammenströmen, schon schwer zu regieren, so scheint die unbekannte Menge, welche die dunkle Hälfte bevölkert, nahezu unberechenbar. Und doch entscheidet die dunkle Welt das Schicksal der hellen.*
>
> Max Reinhardt

1.1 Relevanz und Ziele der Arbeit

Seit den 80er Jahren des vergangenen Jahrhunderts wandelten sich das Verständnis und die Organisation öffentlicher Leistungen nach einer verstärkten Debatte über Qualität und Qualitätssicherung im öffentlichen Sektor (Chaffee & Sherr, 1993; Sowa et al., 2004; Baruch & Ramalho, 2006). In Deutschland führte dies zu Reformen, die unter dem Begriff ‚Neues Steuerungsmodell' bekannt wurden und die Einführung einer marktgesteuerten und kundenorientierten Dienstleistungsproduktion (Bogumil & Naschold, 2000, S. 15) beinhalteten. Die Modernisierung der öffentlichen Verwaltung war das Ziel. Als ein Teil davon ist auch die deutsche Theaterlandschaft betroffen und die Notwendigkeit einer Evaluation ist – sowohl aus Sicht der öffentlichen Theaterförderung im Rahmen der Kulturpolitik (Heinze, 1994, S. 19) als auch aus Sicht eines betriebswirtschaftlich ausgerichteten Theatermanagements (Allmann, 1997) – unumstritten. Sie wird jedoch von zwei Seiten her erheblich erschwert: So sind weder Qualitätskonzepte für den For-Profit-Bereich noch solche für Non-Profit-Organisationen für eine Übertragung auf öffentliche Theater geeignet (Boerner, 2002, S. 59ff.).[1] Dazu kommt, dass weder die Substanz künstlerischer Qualität im Theater als eindeutig definierbar gilt noch Einigkeit über die Kriterien einer Bewertung herrscht (Heinze, 1994, S. 14; Lipp, 1994, S. 557; Hoegl, 1995, S. 24; Allmann, 1997, S. 128). Immerhin existieren Kriterien für die Beschreibung und Bewertung künstlerischer Qualität sowie Operationalisierungsvorschläge, die die künstlerische Qualität jedoch zumeist nicht direkt, sondern über Umwege messen (Ossadnik, 1987, S. 147; KGSt, 1989, S. 29; Schugk, 1996, S. 271; Fabel, 1998, S. 122). In der Praxis des Theatermanagements werden die Kriterien der Publikumsresonanz (Fischer et al., 1994, S. 12; Erhardt, 1994, S. 32) sowie Expertenurteile, z.B. in Form von Theaterkritiken in den Medien, verliehenen Theaterpreisen oder Einladungen zu Gastspielen (Erhardt, 1994, S. 45; Heinze, 1997, S. 280), verwendet. Doch dies wird sowohl von den Theatern selbst als auch in

[1] Als Grund identifiziert Boerner (2002) das Qualitätsverständnis eines Theaters, welches sich hinsichtlich der Kundenorientierung, der Messbarkeit und des Bezugs zu ökonomischen (Formal)zielen der Organisation deutlich von anderen unterscheidet (2002, S. 60ff.).

der Theorie als problematisch angesehen (Boerner, 2002, S. 51; Moser, 2005, S. 11), so dass ein valides und reliables Instrument zur Erfassung künstlerischer Qualität im Theater nach wie vor fehlt. Die vorliegende Arbeit bezieht sich auf Qualität im Musiktheater[1]. Da es wegen seiner Kombination von musikalischen und szenischen bzw. darstellerischen Aspekten als die komplexeste Form darstellender Kunst gilt, erscheint diese Beschränkung auf eine Sparte nicht als inhaltliche Verkürzung.

Das erste Konzept zur künstlerischen Qualität im Musiktheater entwirft Boerner (2002), die zwischen Profilqualität und Interpretationsqualität differenziert. Erste empirische Untersuchungen zur Interpretationsqualität im Musiktheater kommen von Moser (2005), Renz (2006) und Boerner et al. (in Druck).[2] Der erste Teil der vorliegenden Arbeit dient der Validierung des dabei entwickelten Fragebogens zur Erfassung des Rezipientenurteils über die künstlerische Qualität im Musiktheater. Dazu wird der Fragebogen anhand der bisherigen Ergebnisse modifiziert und in einer weiteren Erhebung eingesetzt. Es soll untersucht werden, welche Aspekte einer Aufführung vom Publikum bei der Bewertung unterschieden werden und mit welcher Gewichtung sie in das Gesamturteil über die künstlerische Qualität eingehen. Darüber hinaus stehen Unterschiede zwischen Laien und Experten im Fokus. In einem zweiten Schritt wird die Entstehung des Gesamturteils des Publikums über einen Abend in der Oper untersucht. Es wird vermutet, dass neben der künstlerischen Qualität zusätzliche Faktoren wie bspw. die Erfahrung eines Zuschauers[3], seine Emotionen während der Aufführung sowie situative Rahmenbedingungen die Rezeption einer Musiktheateraufführung beeinflussen (Eversmann, 2004; Boerner & Renz, in Vorbereitung). Dazu werden nach theoretischen Überlegungen zusätzliche Konstrukte in den Fragebogen aufgenommen. Auch interessieren Unterschiede zwischen Laien und Experten hinsichtlich des Gesamturteils, ihnen gilt ein weiterer Teil dieser Arbeit.

Mit der vorliegenden Arbeit soll ein Beitrag zur Entwicklung eines Instrumentes geleistet werden, welches sowohl Inhalt, Struktur und Konsistenz von Urteilen über die künstlerische

[1] Obwohl die Begriffe ‚Oper' und ‚Musiktheater' in der Literatur nicht von allen Autoren synonym verwendet werden (z.B. Mauser, 2002, S. VII), werden sie im Folgenden aus Gründen der Textverständlichkeit gleichgesetzt.

[2] Ursächlich für die Konzentration auf die Interpretationsqualität ist, dass sich öffentliche Musiktheater in Deutschland hinsichtlich der Profilqualität zum Großteil sehr ähnlich sind (Boerner, 2004, S. 427). Dieses zeigt sich bspw. an der geringen Bandbreite der Repertoires, die bei der Mehrheit der Häuser aus einem kleinen Kern aus Werken vorwiegend aus dem 19. Jahrhundert bestehen (Hoegl, 1995, S. 149ff.).

[3] Um die Lesbarkeit zu erleichtern, wird in der vorliegenden Arbeit bei Personenbezeichnungen zumeist die männliche Form verwandt. Diese Bezeichnungen erfassen jedoch selbstverständlich weibliche und männliche Personen.

Qualität einer Opernaufführung zu erfassen in der Lage ist als auch die Entstehung von Gesamturteilen des Publikums über einen Abend in der Oper erklären kann. Zur Anwendung gekommen könnte ein solches Instrument zu einer nachvollziehbaren Vergabe öffentlicher Mittel an die Theater beitragen, die sich aufgrund der zunehmenden Knappheit öffentlicher Mittel[1], eines stetig wachsenden Subventionsbedarfs der Theater[2] und eines gesetzlich nur vage definierten öffentlichen Auftrags[3] schwierig gestaltet. Auch zur betriebswirtschaftlichen Diskussion innerhalb der Theater könnte das neu entwickelte Instrument beitragen. So könnten sich diese verstärkt an den Bedürfnissen des Publikums ausrichten, um Subventionskürzungen aufzufangen und sinkenden Besucherzahlen entgegenzuwirken. Selbst ohne im ständigen Zielkonflikt zwischen dem künstlerischen Anspruch, ökonomischen Notwendigkeiten und Wünschen des Publikums letztere zu sehr in den Mittelpunkt zu stellen und sich damit dem Vorwurf auszusetzen, „nicht mehr anspruchsvoll und avantgardistisch zu sein und lediglich dem Publikumsgeschmack hinterherzulaufen“ (Sievers, 2005, S. 48), kann das zu entwickelnde Instrument verwendet werden. Da nämlich erstmalig auch Aspekte des Publikumsurteils erfasst werden, die über die künstlerische Qualität einer Opernaufführung hinausgehen, könnten durch deren optimale Gestaltung die Besucherzahlen gesteigert werden, ohne den künstlerischen Anspruch zu gefährden. In der Folge könnte die Akzeptanz der Subventionen in der Öffentlichkeit gesteigert werden, wenn nämlich die Theater glaubhaft versichern könnten, sich intensiv für eine Steigerung ihrer Eigeneinnahmen einzusetzen.[4] In

[1] Stiegen die Kulturausgaben der öffentlichen Hand laut Kulturfinanzbericht 2006 des Statistischen Bundesamtes zwischen 1995 und 2003 auf den ersten Blick um 8.1% auf 8.07 Mrd. Euro, zeigt sich ein Rückgang von 0.2%, wenn man die Preisänderungen näherungsweise in Höhe des für das Bruttoinlandsprodukt errechneten Deflators eliminiert. 2003 lagen real gesehen die Ausgaben je Einwohner um 0.2% unter dem Niveau von 1995, und 2005 dürften sie sogar um 2% unter dem Niveau von 1995 gelegen haben (Kulturfinanzbericht 2006, S. 21). Ähnliches gilt, wenn man die Ausgaben für Theater und Musik getrennt betrachtet: 2003 stellten die öffentlichen Haushalte 2.99 Mrd. Euro aus allgemeinen Haushaltsmitteln für diesen Bereich zur Verfügung, was 37% aller im Jahr 2003 von Bund, Ländern und Gemeinden getätigten Kulturausgaben entspricht. 2001 waren es 3.1% mehr gewesen. Für 2005 betrugen die öffentlichen Ausgaben für Theater und Musik im Jahr 2005 voraussichtlich 2.98 Mrd. Euro (Kulturfinanzbericht 2006, S. 62).

[2] Dieser liegt an außerhalb des Einflussbereiches öffentlicher Theater stehenden Kostensteigerungen, bspw. durch die Erhöhung der Tarifverträge. Im Gegensatz zu Industriebetrieben sind Theater als „archaische Produktionsform“ (Fabel, 1998, S. 140) nicht in der Lage, die Verteuerung des Faktors Arbeit durch technologische Entwicklungen und Rationalisierungsmassnahmen auszugleichen. Dieser Zusammenhang wurde erstmalig von den US-amerikanischen Ökonomen Baumol und Bowen erläutert und als *cost disease of the performing arts* bezeichnet (Baumol & Bowen, 1966, S. 161ff.).

[3] Im Unterschied zum Auftrag anderer öffentlicher Unternehmen ist die Kulturpolitik in Deutschland eine freiwillige Aufgabe der Rechtsträger (Heinrichs, 1995, S. 302), so dass die jeweiligen politischen Entscheidungsträger des Rechtsträgers von Theatern über die konkrete Definition der öffentlichen Aufgabe entscheiden (Göschel, 1997).

[4] Nach der Statistik des Deutschen Bühnenvereins lag der Kostendeckungsgrad im Jahr 2004/05 bei durchschnittlich 17% des Gesamtbudgets der öffentlichen Theater Deutschlands, wobei die Einnahmen aus dem Kartenverkauf die wichtigste Quelle der Eigeneinnahmen darstellte.

der Summe würden diese Maßnahmen die Erfüllung der gesellschaftlichen Funktionen öffentlicher Theater[1] verbessern und längerfristig sicherstellen helfen.

1.2 Vorgehen

In Kapitel 2 werden die Untersuchungsfragen vorgestellt und ein Überblick über den Stand der Forschung gegeben. Dazu werden zunächst das bislang einzige analytische Modell zur Erfassung künstlerischer Qualität im Musiktheater (Boerner, 2002) erläutert und Untersuchungen zu seiner empirischen Überprüfung präsentiert. Anschließend wird Literatur zur Erklärung individueller Urteile über einen Abend im Konzert sowie im Sprechtheater vorgestellt. Das Kapitel endet mit der Erläuterung der im Rahmen der vorliegenden Arbeit hergeleiteten Determinanten des Gesamturteils des Publikums über einen Abend in der Oper (‚Gesamterlebnis Opernabend')

Es folgen die Darstellungen des Instrumentes der vorliegenden Untersuchung sowie der verwendeten Methode in Kapitel 3.

In Kapitel 4 werden vorbereitenden Analysen dargestellt, die der Klärung der Untersuchungsfragen dienen. Dazu gehören sowohl deskriptive Statistiken als auch Untersuchungen zu Reliabilität und Validität des entwickelten Fragebogens.

Kapitel 5 beschäftigt sich mit der Analyse und Interpretation der erhobenen Daten. Es werden die Untersuchungsfragen dieser Arbeit erörtert.

Die Schlussbetrachtung in Kapitel 6 fasst die Ergebnisse der Datenauswertung zusammen. Abschließend werden die Untersuchungsanordnung kritisch diskutiert und mögliche weitere Forschungsschritte erörtert.

[1] Dazu gehören u.a. die Förderung der Künste und Künstler, die kulturelle Bildung und Kulturpädagogik, die Pflege und Erforschung des öffentlichen Erbes sowie die Entwicklung der soziokulturellen Gestalt- und Lebensqualität der Städte (Sievers, 2005, S. 46; für eine ausführliche Diskussion vgl. Boerner, 2002, S. 22 ff.).

2. Fragestellung der Arbeit, Stand der Forschung und eigene Überlegungen

2.1 Untersuchungsfragen

Folgende Fragestellungen sollen im Verlauf dieser Arbeit untersucht werden:

Untersuchungsfrage 1: Was ist Gegenstand der Bewertung künstlerischer Qualität im Musiktheater?

a) Welche musikalischen und szenischen Aspekte einer Opernaufführung beachten und unterscheiden die Rezipienten bei ihren Urteilen in welchem Differenzierungsgrad?

b) Mit welcher Gewichtung gehen die unterschiedlichen Teilaspekte in das Gesamturteil der Rezipienten über die künstlerische Qualität einer Opernaufführung ein?

Untersuchungsfrage 2: Wie wird die Aufführungsqualität von verschiedenen Beurteilergruppen, d.h. von Laien und Experten, bewertet?

a) Welche Unterschiede bestehen zwischen den beiden Beurteilergruppen in Bezug auf Differenziertheit und Gewichtung der Teilaspekte?

b) Wird die Qualität der Teilaspekte einer Aufführung von den beiden Beurteilergruppen unterschiedlich beurteilt?

c) Wie hoch ist die Übereinstimmung (Interrater-Reliabilität) der Urteile innerhalb der beiden Beurteilergruppen?

Untersuchungsfrage 3: Welche Determinanten beeinflussen das Gesamturteil des Publikums über einen Abend in der Oper (‚Gesamterlebnis Opernabend')? Welche Wirkungen haben Aspekte über die Bewertung der künstlerischen Qualität hinaus?

Untersuchungsfrage 4: Wie wird das ‚Gesamterlebnis Opernabend' von verschiedenen Beurteilergruppen, d.h. von Laien und Experten, bewertet?

a) Werden die Teilaspekte des Gesamterlebnisses von den beiden Beurteilergruppen unterschiedlich bewertet?

b) Wie hoch ist die Übereinstimmung (Interrater-Reliabilität) der Urteile innerhalb der beiden Beurteilergruppen?

c) Welche Unterschiede bestehen zwischen den beiden Beurteilergruppen in Bezug auf die Gewichtung der Teilaspekte?

Untersuchungsfrage 1 beschäftigt sich mit dem Inhalt und der Struktur des Publikumsurteils über die künstlerische Qualität einer Opernaufführung. Wahrnehmungs- und Urteilsprozesse werden in dieser Arbeit entsprechend der Konzeption von Upmeyer (1985) verstanden, der diese auf den drei Ebenen ‚Stimulusebene', ‚Ebene der internen Repräsentation' und ‚Antwortebene' darstellt. *Wahrnehmung* wird dabei als Transformation von der Stimulusebene auf die Ebene der Repräsentation verstanden, wohingegen die *Urteilsbildung* als Transformation von der internen Repräsentationsebene auf die Antwortebene aufgefasst wird. Als *Urteil* gilt somit das Ergebnis dieser Transformation (Niketta, 1993, S. 330).[1] Im Rahmen der Untersuchungsfrage 1a wird geklärt werden, welche Teilaspekte der musikalischen Dimension (z.B. Orchester, Solisten, Chor) und der szenischen Dimension (z.B. Kostüme, Bühnenbild) einer Aufführung im Musiktheater von den Zuschauern in welchem Differenzierungsgrad unterschieden werden. Untersuchungsfrage 1b widmet sich daran anschließend der Frage, mit welcher Gewichtung die unterschiedlichen Teilaspekte in das Urteil der Zuschauer über die künstlerische Qualität eingehen. So ist bspw. zu klären, ob die Aspekte der musikalischen und der szenischen Dimension gleich stark gewichtet sind.

Im Rahmen von Untersuchungsfrage 2 wird untersucht werden, ob verschiedene Beurteilergruppen die künstlerische Qualität einer Opernaufführung unterschiedlich bewerten. Da die musikalische Vorbildung in der musikpsychologischen Forschung als der Faktor gesehen wird, der das Urteil über Musik am nachhaltigsten beeinflusst (Morguet & Moser-Hauck, 1991, S. 189) und ähnliche Effekte auch im Musiktheater zu erwarten sind (Boerner & Renz, in Vorbereitung), wird in dieser Arbeit zwischen Laien und Experten differenziert werden. So versucht Untersuchungsfrage 2a zu klären, ob sich Laien und Experten in Hinblick auf die Differenziertheit und Gewichtung der verschiedenen Teilaspekte unterscheiden. Untersuchungsfrage 2b widmet sich der Frage, ob die Qualität der Teilaspekte einer Aufführung von den beiden Beurteilergruppen unterschiedlich beurteilt wird, wohingegen in Untersuchungsfrage 2c geklärt wird, wie hoch die Übereinstimmung (Interrater-Reliabilität) der Urteile über die künstlerische Qualität innerhalb der beiden Beurteilergruppen ist.

Im weiteren Verlauf der vorliegenden Arbeit steht das Publikumsurteil über den gesamten Abend in der Oper im Mittelpunkt (‚Gesamterlebnis Opernabend'). Es wird vermutet, dass

[1] Wenngleich in der Literatur nicht immer gleichgesetzt, werden die Begriffe ‚Urteil', ‚Beurteilung', ‚Bewertung' und ‚Rezeption' des Publikums als seine Reaktion auf Gesehenes bzw. Gehörtes im Folgenden synonym verwendet (für eine ausführliche Diskussion vgl. Schulten, 1990).

Aspekte über die Bewertung der künstlerischen Qualität hinaus in das Gesamturteil einfließen (Untersuchungsfrage 3).

In Untersuchungsfrage 4 interessieren wiederum Unterschiede zwischen Laien und Experten. Innerhalb Untersuchungsfrage 4a soll geklärt werden, ob die Teilaspekte des Gesamterlebnisses von Laien und Experten unterschiedlich beurteilt werden. Der Höhe der Übereinstimmung (Interrater-Reliabilität) der Urteile innerhalb der beiden Beurteilergruppen widmet sich Untersuchungsfrage 4b, wohingegen sich Untersuchungsfrage 4c abschließend mit Unterschieden zwischen den beiden Beurteilergruppen in Bezug auf die Gewichtung der Teilaspekte beschäftigt.[1]

2.2 Stand der Forschung

Im Folgenden wird der Stand der Forschung präsentiert. Auf die Darstellung des analytischen Modell Boerners zur Konzeption der Aufführungsqualität in der Oper (Boerner, 2002) in Abschnitt 2.2.1 folgen empirische Untersuchungen auf der Basis des Modells (Abschnitt 2.2.2) sowie Literatur zur Erklärung individueller Reaktionen im Musiktheater (Abschnitt 2.2.3).

2.2.1 Das analytische Modell zur künstlerischen Qualität im Musiktheater

In ihrer Arbeit zu Führungsbesonderheiten im Musiktheater entwickelt Boerner (2002) das erste analytische Modell zur künstlerischen Qualität im Musiktheater, das der Präzisierung des angestrebten Führungserfolges dient. Dabei interessieren sie eine systematische Beschreibung bzw. Definition der künstlerischen Qualität sowie die Frage der Messbarkeit derselben.

Boerner unterteilt das Konstrukt ‚künstlerische Qualität im Musiktheater' in zwei Aspekte: Profilqualität und Interpretationsqualität. Unter Profilqualität versteht sie dabei „das Leistungsprogramm eines Hauses" (Boerner, 2002, S. 133), das sich im Spielplan konkretisiert und sämtliche angebotenen Veranstaltungen umfasst. Die Kriterien für die Beschreibung der Profilqualität eines öffentlichen Theaters, die die Autorin aus dessen Kulturauftrag ableitet, sind folgende: Vielfalt vs. Spezialisierung, Konformität vs. Originalität sowie Moderne vs. Traditionalität des Spielplans. Für die vorliegende Arbeit wesentlich

[1] Da in der Literatur noch kaum Erkenntnisse hinsichtlich der in dieser Arbeit gestellten Untersuchungsfragen bestehen, wird auf die Formulierung von Hypothesen bewusst verzichtet. Die wenigen vorhandenen Ergebnisse werden jedoch bei der Datenauswertung in Kapitel 5 vergleichend berücksichtigt.

relevanter ist jedoch der zweite Aspekt künstlerischer Qualität im Musiktheater: die von der Profilqualität zumindest teilweise unabhängige Interpretationsqualität. Diese wird als „Ausführung des Leistungsprogramms“ (Boerner, 2002, S. 64) definiert und umfasst die Produktion (Inszenierung und musikalische Konzeption) sowie die Aufführung eines Werkes. Da die Oper ein mehrdimensionales Werk darstellt (Koebner, 1993, S. 193; Pahlen, 1981, S. 11), unterteilt Boerner auch die Interpretationsqualität in die Dimensionen ‚Musik’ und ‚Szene’. Jede Dimension wiederum wird in einem weiteren Schritt in Potential- und Ergebnisfaktoren differenziert. Potentialfaktoren werden als Teilqualitäten definiert, „die in ihrer Gesamtheit die Interpretationsqualität im Musiktheater ausmachen“ (Boerner, 2002, S. 79). Ergebnisfaktoren hingegen „differenzieren die pro Potentialfaktor angestrebte Interpretationsqualität (z.B. ein bestimmtes Klangbild, eine bestimmte Atmosphäre) weiter aus“ (Boerner, 2002, S. 79). Tabelle 1 dient zur Veranschaulichung.

Tabelle 1: Kombination von Potential- und Ergebnisfaktoren der musikalischen und der szenischen Dimension der Interpretationsqualität im Musiktheater

Musikalische Dimension			
	Potentialfaktoren		
Ergebnisfaktoren			
	Orchesterqualität	Chorqualität	Qualität des Sologesangs
Klang			
Tempo/ Rhythmus			
Szenische Dimension			
	Potentialfaktoren		
Ergebnisfaktoren			
	schauspielerische Qualität	Ausstattungsqualität	
Handlung			
Ort			
Zeit			
Figuren			
Atmosphäre			
Genre			

aus: Boerner, 2002, S. 80

Eine Verfeinerung dieses zweidimensionalen Beschreibungsschemas strebt Boerner durch eine Ausdifferenzierung der Potential- und Ergebnisfaktoren an (Boerner, 2002, S. 81 ff.). Da die Interpretationsqualität im Rahmen dieser Arbeit lediglich einen Teil der künstlerischen Qualität im Musiktheater darstellt, wird das bereits beschriebene Differenzierungsniveau der Potential- und Ergebnisfaktoren hier als ausreichend betrachtet.

Die Qualität der einzelnen Potential- und Ergebnisfaktoren sowie ihrer jeweiligen Teilfaktoren ist nach Boerner jedoch nicht allein ausschlaggebend für die Interpretationsqualität einer Opernaufführung. Ebenfalls entscheidend ist die Stimmigkeit der Faktoren untereinander, welche als die „Integration aller Teilmomente eines Kunstwerkes zu einem Ganzen“ (Boerner, 2002, S. 92) beschrieben und als zentrales Qualitätskriterium gesehen wird. Neben der Interpretationsqualität ist auch das Kriterium der Stimmigkeit einer Aufführung, das in Anlehnung an die kontingenztheoretische Forschung als ‚fit’ bezeichnet wird, bis zu diesem Zeitpunkt nicht systematisch untersucht worden, weswegen Boerner einen eigenen Vorschlag zur Präzisierung präsentiert. Der fit wird darin in einem hierarchischen Modell konkretisiert und bezieht sich jeweils auf die oben erläuterten Ergebnisfaktoren der Interpretationsqualität.

An erster Stelle ist die Stimmigkeit zwischen der musikalischen und der szenischen Dimension bedeutsam, die als „fit erster Ordnung“ oder „fit 1“ bezeichnet wird. Boerner erwähnt als Beispiel die Abstimmung zwischen dem in der musikalischen Dimension erzeugten Klangbild und der Atmosphäre, die die szenische Dimension vermittelt (Boerner, 2002, S. 93). Auch innerhalb der jeweiligen Dimensionen ist eine Stimmigkeit anzustreben, die „fit zweiter Ordnung“ bzw. „fit 2“ genannt wird. So sind „innerhalb der musikalischen Dimension […] die Potentialfaktoren (Orchesterqualität, Chorqualität und die Qualität des Sologesangs) in Bezug auf die Ergebnisfaktoren (Klang und Tempo/ Rhythmus) aufeinander abzustimmen“ (Boerner, 2002, S. 93). Selbiges gilt für die Potentialfaktoren der szenischen Dimension, innerhalb derer die Abstimmung der schauspielerische Qualität und der Ausstattungsqualität bspw. in Bezug auf die Handlung gegeben sein muss. Drittens ist die Stimmigkeit innerhalb der jeweiligen Potentialfaktoren bedeutsam, die als „fit dritter Ordnung“ oder „fit 3“ bezeichnet wird. Hierunter fallen bspw. die Synchronisation der Orchestermusiker in Bezug auf ihre Einsätze sowie die Ausrichtung von Kostümbild und Bühnenraum auf die szenische Gesamtkonzeption. Insgesamt ist das Stimmigkeitsmerkmal also als „ein Ergebnisfaktor einzuordnen, der das Zusammenspiel der Teilfaktoren, der Potentialfaktoren und der Dimension beschreibt und daher eine Art Meta-Kriterium der Interpretationsqualität darstellt“ (Boerner, 2002, S. 93).[1]

Das von Boerner entwickelte Konzept künstlerischer Qualität im Musiktheater ist dabei hierarchisch aufgebaut, so dass die Stimmigkeit auf jeder Ebene erneut hergestellt werden

[1] Ergänzend sei angemerkt, dass Stimmigkeit jedoch keineswegs eine vollständige Parallelisierung voraussetzt, die eher als komisch gilt und daher als „Mickey-Mousing“ (Kühn, 1982, S. 12) bezeichnet wird.

muss und jeder untergeordnete fit eine notwendige, jedoch nicht hinreichende Bedingung für den fit auf der nächst höheren Ebene darstellt (Boerner, 2002, S. 97). Aus diesem Grund sind die drei genannten Ebenen „so miteinander verbunden, dass Entscheidungen auf der übergeordneten Ebene jeweils Entscheidungs- und Handlungsspielräume auf den nachgeordneten Ebenen einschränken“ (Boerner, 2002, S. 97). Die drei fit-Kriterien sind also nicht unabhängig voneinander, wobei die Abhängigkeit je nach Merkmal unterschiedlich stark sein kann. Tabelle 2 dient wiederum zur Veranschaulichung.

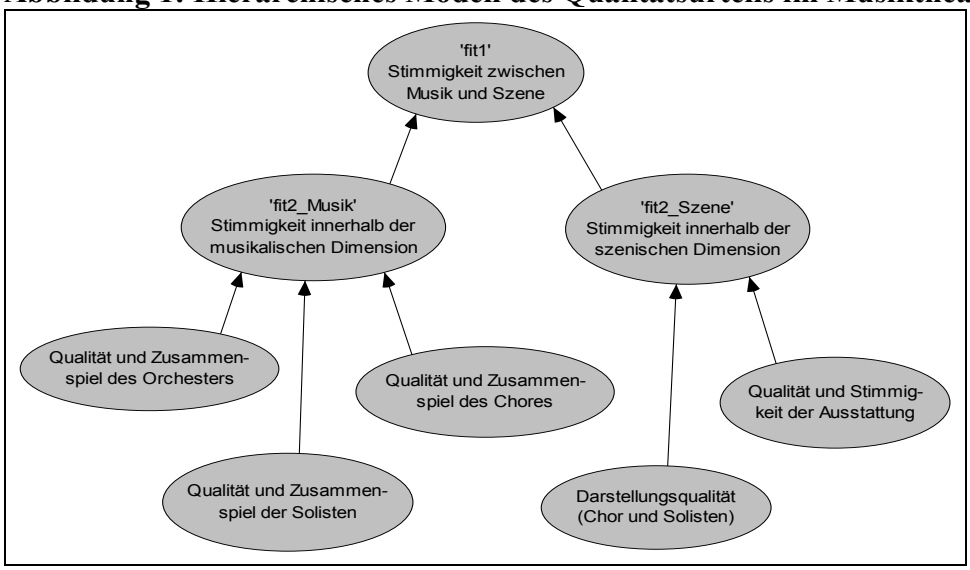
Tabelle 2: Hierarchisches fit-Modell der Interpretationsqualität im Musiktheater

fit 1 ***Stimmigkeit der Dimensionen in Bezug auf die Ergebnisfaktoren***				
Musikalische Dimension			**Szenische Dimension**	
fit 2 ***Stimmigkeit der Potentialfaktoren in Bezug auf die Ergebnisfaktoren der musikalischen Dimension***			**fit 2** ***Stimmigkeit der Potentialfaktoren in Bezug auf die Ergebnisfaktoren der szenischen Dimension***	
Orchester-qualität (Instrumente, Stimmen, Pulte)	Chorqualität (Stimmen)	Qualität des Sologesangs (Stimmen)	schauspielerische Qualität (Tätigkeit, Erscheinung)	Ausstattungsqualität (Kostümbild, Raumkonzeption, Bühnenraum)
fit 3	**fit 3**	**fit 3**	**fit 3**	**fit 3**
Stimmigkeit der Teilfaktoren				

aus: Boerner, 2002, S.92

Eine grafische Darstellung der hierarchischen Beziehungen der Dimensionen, Potentialfaktoren und Teilfaktoren zwischen den verschiedenen Modellebenen bietet Abbildung 1.

Abbildung 1: Hierarchisches Modell des Qualitätsurteils im Musiktheater



nach: Renz, 2006, S. 9

Wie Abbildung 1 zu entnehmen ist, besteht die dritte Ebene des analytischen Modells sowohl aus Potentialfaktoren als auch aus dem Zusammenspiel innerhalb der jeweiligen Potentialfaktoren. Zehn Konstrukte lassen sich differenzieren: Diese sind die musikalische Leistung der Solisten, des Orchesters und des Chores, die szenische Leistung der Solisten und des Chores, die Qualität der Ausstattung sowie das Zusammenspiel innerhalb der Potentialfaktoren (Zusammenspiel der Solisten, innerhalb des Orchesters sowie innerhalb des Chores, Stimmigkeit der szenischen Aspekte der dritten Ebene). Wie beschrieben beeinflussen diese die zweite Modellebene, die wiederum auf die erste Ebene Einfluss nimmt.

2.2.2 Empirische Überprüfung des analytischen Modells

Nach der theoretischen Erfassung sämtlicher Aspekte, die Gegenstand bei der Beurteilung der künstlerischen Qualität in der Oper sein können, muss empirisch überprüft werden, welche Teilaspekte in welchem Differenzierungsgrad vom Publikum tatsächlich erfasst werden. Eine empirische Überprüfung des Modells führt Moser (2005) durch, der zur Entwicklung eines Fragebogens zur Erfassung des Publikumsurteils über die künstlerische Qualität im Musiktheater in einem ersten Schritt Aufführungskritiken überregionaler Zeitungen analysiert (2005, S. 14ff.). Ziel dabei sind Erkenntnisse über die Kriterien der Bewertung einer Opernaufführung durch professionelle Kritiker. In einem nächsten Schritt gleicht Moser die Ergebnisse der Inhaltsanalyse mit dem analytischen Modell Boerners (2002) ab (Moser, 2005, S. 22ff.) und führt eine Häufigkeitsanalyse durch, um die Gewichtung der einzelnen Teilkonstrukte zu bestimmen (2005, S. 28).[1] Aus den so erhaltenen Ergebnissen entwickelt Moser einen Fragebogen zur Messung der künstlerischen Qualität in der Oper, in dem die Konstrukte des analytischen Modells, die allesamt durch die Inhaltsanalyse der Aufführungskritiken bestätigt werden[2], durch verschiedene Items operationalisiert werden (2005, S. 32ff.). Zur Erfassung der Gewichtung der Teilaspekte auf das Gesamturteil über die künstlerische Qualität operationalisiert er zusätzlich das Konstrukt ‚Gesamturteil' als abhängige Variable. Zum Einsatz kommt der Fragebogen im Rahmen eines Pretests an Studierenden der Universität Konstanz, denen Ausschnitte von Opernaufführungen auf Video gezeigt werden. Aus den erhobenen Daten zieht Moser abschließend erste empirische Erkenntnisse hinsichtlich der Urteilsfindung zur künstlerischen Qualität in der Oper: So zeigt

[1] In den Aufführungskritiken zeigt sich, dass die Kriterien der musikalischen Dimension nahezu doppelt so oft verwendet werden wie die der szenischen Dimension. Darüber hinaus wird die musikalische Dimension erheblich stärker differenziert. Moser zieht den Schluss, dass die Kritiker die musikalische Komponente einer Opernaufführung für wesentlich wichtiger erachteten als die szenische Komponente (2005, S. 29).

[2] Zusätzlich zu den vom analytischen Modell verwendeten entdeckt Moser (2005) in der Empirie weitere Kriterien: Diese sind der Dirigent, die Werktreue, der Interpretationsspielraum für die Phantasie der Zuschauer, die Originalität beim Orchester sowie die Wirkung der Aufführung auf die Zuschauer (2005, S. 23).

sich, dass zwar die ‚physisch' unterscheidbaren Konstrukte ‚Orchester', ‚Chor', ‚Solisten' und ‚Ausstattung' unterschieden werden, jedoch nicht deren Unterkonstrukte (z.B. musikalische Dimension, szenische Dimension). Als Konstrukt mit der stärksten Gewichtung geht die ‚Ausstattung' in das Gesamturteil der Befragten ein (2005, S. 70ff.).

Im Rahmen ihrer Diplomarbeit 'Evaluation in Non-Profit Organisationen – Eine empirische Untersuchung zur Rezeption der Aufführungsqualität im Musiktheater' entwickelt Renz (2006) den Fragebogen aus den Ergebnissen bei Moser (2005) sowie eigenen Faktoren- und Reliabilitätsanalysen weiter. Sie versucht genauer zu klären, welche Aspekte einer Opernaufführung vom Publikum bei der Beurteilung der künstlerischen Qualität unterschieden werden und mit welcher Gewichtung diese jeweils in das Gesamturteil eingehen. Außerdem interessieren sie eventuelle Unterschiede in den Bewertungen verschiedener Beurteilergruppen, d.h. Laien und Experten, sowie die Hierarchie der Urteilsstruktur (2006, S. 4ff.). Der von Renz modifizierte Fragebogen kommt Anfang 2006 ebenfalls an Studierenden der Universität Konstanz zum Einsatz. Zur besseren Vergleichbarkeit wird dabei dasselbe Stimulusmaterial, d.h. Ausschnitte derselben Opernaufführungen, verwendet wie bei der Untersuchung Mosers (2005). Renz bestätigt das Ergebnis Mosers, nach dem das Publikum in seinem Urteil über die künstlerische Qualität einer Opernaufführung zwischen unterschiedlichen Teilaspekten differenziert. Sie zeigt außerdem, dass das Publikumsurteil – wie im analytischen Modell Boerners (2002) angenommen – hierarchisch aufgebaut ist und dass die Teilaspekte unterschiedlich gewichtet in das Gesamturteil eingehen. Unterschiede in den Bewertungen von Laien und Experten ergeben sich hinsichtlich des Inhalts, der Struktur sowie der Konsistenz der Qualitätsurteile. Die interindividuelle Übereinstimmung der Urteile ist in beiden Gruppen tendenziell gering (Renz, 2006, S. 91).

Zum ersten Mal zur Befragung eines ‚echten' Publikums wird der Fragebogen im Mai 2006 von Boerner et al. (in Druck) verwendet. Dabei werden 145 Besucher einer Aufführung der Oper „La forza del destino" von Giuseppe Verdi in Köln befragt.

Im Gegensatz zu den vorgestellten Untersuchungen, die sich allein der künstlerischen Qualität einer Opernaufführung widmen, interessiert in der vorliegenden Arbeit das Gesamturteil des Publikums über einen Abend in der Oper (‚Gesamterlebnis Opernabend'). Da vermutet wird, dass die Qualität der Kunst nicht das einzig ausschlaggebende Kriterium für die

Publikumsreaktion darstellt, wie lange Zeit angenommen wurde (u.a. Thompson, 2007, S. 20f.), soll im Folgenden jene Literatur vorgestellt werden, die bei ihren Erklärungsversuchen nicht ausschließlich die Qualität der Kunst betrachtet.

2.2.3 Erklärungen individueller Reaktionen im Musiktheater

Nur wenig Literatur beschäftigt sich speziell mit den Determinanten des Gesamturteils des Publikums über einen Abend in der Oper (‚Gesamterlebnis Opernabend'), welches in der vorliegenden Arbeit im Mittelpunkt steht (Untersuchungsfragen 3 und 4). Allein Behr (1983) unternimmt den Versuch, eine Opernaufführung als Ganzes zu sehen, und sucht nach deren Effekten auf das Publikum (n = 1.033, sieben Aufführungen). Es gelingt ihm, sechs Wirkungen einer Musiktheateraufführung zu isolieren: eine emotionalisierende, eine archetypisierende, eine ideologisierende, eine kritisch-rationalisierende, eine stabilisierende sowie eine ästhetisch-sensibilisierende Wirkung, wobei erstgenannte die größte Zustimmung erhält (Behr, 1983, S. 165). Wie sich diese jedoch auf das Gesamturteil des Publikums über eine Opernaufführung auswirken, bleibt unklar. Um für die vorliegende Arbeit einen breiteren theoretischen Hintergrund zu erhalten, erscheint daher auch der Einbezug von Literatur erforderlich, die sich nicht explizit mit der Oper beschäftigt. Sowohl Literatur zur Erklärung individueller Urteile über einen Abend im Konzert (Abschnitt 2.2.3.1) als auch über einen Abend im Sprechtheater (Abschnitt 2.2.3.2) bieten sich an.

2.2.3.1 Erklärungen individueller Urteile über einen Abend im Konzert

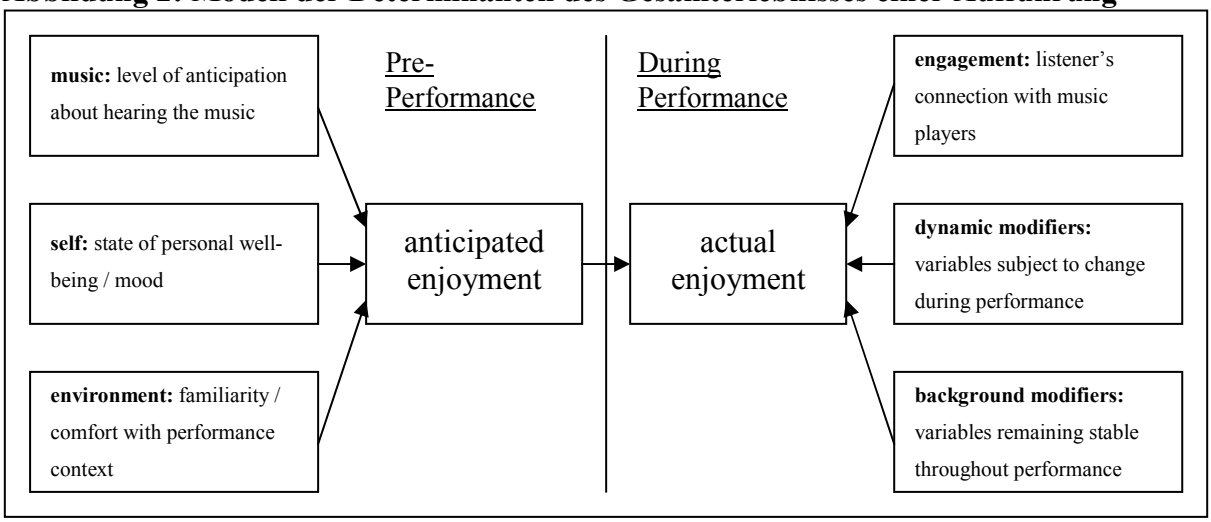
Zur Erklärung individueller Urteile über einen Abend im Konzert werden aus Mangel an Arbeiten, die sich explizit damit beschäftigen, zunächst Untersuchungen vorgestellt, die das Urteil des Publikums über die musikalische Aufführung als abhängige Variable wählen (Abschnitt 2.2.3.1.1). Da in der vorliegenden Arbeit jedoch das Publikumsurteil über den gesamten Abend im Fokus steht, folgen die Darstellung von Arbeiten zur individuellen Reaktion auf Merkmale des aufgeführten Werkes selbst (Abschnitt 2.2.3.1.2) sowie zu Subjekt-Variablen des Zuhörers, wobei insbesondere auf seine emotionale Reaktion auf die musikalische Aufführung eingegangen wird (Abschnitt 2.2.3.1.3). Der Abschnitt schließt mit Arbeiten zu den situativen Bedingungen einer musikalischen Aufführung, bei denen in der vorliegenden Arbeit ebenfalls ein Einfluss auf das Gesamturteil des Publikums über einen Abend im Konzert angenommen wird (Abschnitt 2.2.3.1.4).

2.2.3.1.1 Das Gesamterlebnis ‚Konzertaufführung'

Thompson (2006) interessiert in erster Linie, ob sich Laien und Experten in ihren Urteilen über die künstlerische Qualität einer musikalischen Aufführung unterscheiden, ob Vertrautheit *(prior familiarity)* mit und Gefallen *(liking)* am Stück das Gesamterlebnis beeinflussen, sowie die Beziehung zwischen der wahrgenommenen künstlerischen Qualität einer Aufführung und dem Gesamterlebnis. Dies testet er anhand eines Fragebogens, der in der Pause eines Konzertes des Royal College of Music (London) verteilt wurde (n = 91). Es ergeben sich folgende Ergebnisse, die aufgrund methodischer Schwächen der Studie jedoch vorsichtig zu interpretieren sind: Experten, d.h. Zuhörer, die Musik studiert hatten oder zu jener Zeit studierten, bewerten die künstlerische Qualität niedriger als Laien, wohingegen die restlichen Dimensionen sehr ähnlich bewertet werden. Gefallen am Stück korreliert positiv (r = .67 bzw. r = .59)[1] mit der Bewertung des Gesamterlebnisses; Vertrautheit mit dem Stück hingegen hat keinen Einfluss (r = .09 bzw. r = .05). Als Ergebnis der dritten Untersuchungsfrage Thompsons zeigt sich, dass die emotionale Reaktion *(affective response)* auf das Konzert ein deutlich größeres Gewicht auf das Urteil über das Gesamterlebnis hat als die Bewertung der künstlerischen Qualität *(evaluative response)* (Thompson, 2006). Da die Regressionsanalyse jedoch 60% der Gesamtvarianz unerklärt lässt und daher zu vermuten ist, dass eine große Anzahl der Determinanten des Gesamterlebnisses nicht im Modell enthalten ist, wählt Thompson anschließend bei einer – von der Methode her vergleichbaren – Untersuchung einen anderen Ansatz (2007, n = 264). Dabei konzentriert er sich auf die emotionale Reaktion der Zuhörer auf eine Konzertaufführung, also „the affective experience of the regular concertgoer – what will be termed […] their *enjoyment* of the performance as a whole" (Thompson, 2007, S. 20). 22 Variablen, die er in Expertengesprächen ermittelt, teilt er in zwei Kategorien: ‚Pre-Performance' und ‚During Performance', wobei erstere die Erwartungen des Zuhörers an die Aufführung *(anticipated enjoyment)* bestimmt, wohingegen letztere die tatsächliche Erfahrung *(actual enjoyment)* beeinflusst. Er möchte zeigen, dass „the apparently multitudinous factors influencing affective responses to a performance occasion might, in principle, be described according to a simpler set of underlying components" (Thompson, 2007, S. 33). Dies gelingt ihm mit Hilfe einer Hauptkomponentenanalyse, in der er insgesamt sechs Faktoren extrahieren kann (siehe Abbildung 2).

[1] Zwei Stücke, die jeweils unterschiedlich bekannt waren, wurden vom Orchester gespielt. Diese sollten von den Teilnehmern jeweils getrennt bewertet werden.

Abbildung 2: Modell der Determinanten des Gesamterlebnisses einer Aufführung



aus: Thompson, 2007, S. 31

Von den sechs gezogenen Faktoren gehören drei zur ‚Pre-Performance'-Kategorie, die die Erwartungen beeinflusst. Dazu gehört die Musik selbst *(music)*, die insbesondere die Vorfreude auf das aufgeführte Stück umfasst. Ein beispielhaftes Item lautet ‚I know that I like the music being played'. Der zweite Faktor beinhaltet den persönlichen Zustand des Zuhörers *(self)*, wozu u.a. seine Stimmung gehört. Als dritten Faktor identifiziert Thompson die Vertrautheit des Zuhörers mit dem Kontext bzw. der Umgebung der Aufführung (*environment*, z.B. ‚I am in a familiar venue'). Die verbleibenden drei Faktoren gehören zur ‚During Performance'-Kategorie, die die tatsächliche Wahrnehmung der Aufführung beeinflusst. Als ersten nennt Thompson den Faktor *engagement*, der die Aufmerksamkeit des Zuhörers sowie das gezeigte Engagement der Musiker umfasst. Unter *dynamic modifiers* versteht Thompson Variablen mit Einfluss darauf, wie der Zuhörer die Musiker wahrnimmt bzw. wie konzentriert er ist.[1] Beispielhafte Items lauten ‚The performers play a lot of wrong notes' sowie ‚There are unwelcome distractions (e.g. audience or traffic noise)'. Den sechsten Faktor bilden die *background modifiers*, die ebenfalls Variablen umfassen, die die Wahrnehmung des Zuhörers beeinflussen. Im Gegensatz zu den *dynamic modifiers* sind dies jedoch ausschließlich solche, die während der Aufführung keinen Änderungen unterliegen (z.B. ‚The performers are not appropriately dressed') (Thompson, 2007, S. 26f.).

Da zu vermuten ist, dass weitere Variablen bestehen, die das individuelle Urteil über einen Abend im Konzert beeinflussen, wird ein kurzer Überblick über die Rezeptionsforschung folgen, die sich als interdisziplinäres Forschungsgebiet mit der Wahrnehmung, dem Erleben sowie der Wirkung von Musik beschäftigt (Gembris, 1996). Im Unterschied zu den

[1] Die genaue Abgrenzung zum Faktor *engagement* bleibt unklar.

Untersuchungen Thompsons (2006, 2007) stehen hier jedoch nicht Konzertaufführungen als Ganzes, sondern Teilaspekte im Fokus.

2.2.3.1.2 Der Stimulus ‚aufgeführtes Werk'

„New Experimental Aesthetics"

Einen Versuch, individuelle Reaktionen auf ein Werk mit seinen Merkmalen zu erklären, stellt die von Berlyne angestoßene Tradition der „New Experimental Aesthetics" dar, die einen Zusammenhang zwischen sog. ‚kollativen Variablen' und dem musikalischen Urteil der Befragten postuliert (Berlyne, u.a. 1960, 71, 74a, 74b). Unter kollative Variablen fallen dabei Neugier auslösende Variablen wie Komplexität und Neuartigkeit. Diese sind ebenfalls unter der Bezeichnung ‚Variablen des Vergleichs' bekannt, da sie immer in Beziehung zu anderen gesetzt werden: So wird Neuartigkeit abgegrenzt von vertrauten Erscheinungen, Komplexität ist die Qualität einer gegliederten Struktur (Berlyne, 1974, S. 10). Bestimmende Annahme der „New Experimental Aesthetics" ist dabei, dass die Beziehungen zwischen der Ausprägung der kollativen Variablen und dem ästhetischen Urteil kurvilinear, d.h. umgekehrt U-förmig, verlaufen. Begründet wird dies durch das – in der Psychologie häufig verwendete und nicht weiter begründbare – Homöostase-Prinzip, welches „lebende Organismen zur Vermeidung sowohl von sehr monotonen als auch konfliktinduzierenden oder allzu intensiven Reizungen veranlaßt" (Raab, 1981, S. 264). Der Prozess, der dem von Berlyne beschriebenen Zusammenhang zugrunde liegt, ist das ‚arousal-Konzept', welches eine Wechselwirkung von kognitiver Belastung *(arousal)* und der Ausprägung der jeweiligen kollativen Variablen auf das Urteil der Befragten konstatiert.

Komplexität

Berlyne gelingt es verschiedentlich (u.a. 1970, 1974b), den postulierten Zusammenhang empirisch zu zeigen, wobei anzumerken ist, dass seine Methoden den heutigen Standards nicht mehr entsprechen. Neben Berlyne führt eine Vielzahl von Forschern eigene empirische Untersuchungen durch, deren Ergebnisse jedoch zum Teil widersprüchlich sind[1]: Bestätigung für Berlynes Annahmen finden sich in Studien mit synthetischen Materialien, d.h. künstlich hergestellten Reizmustern, bei denen die kollativen Eigenschaften objektiv bestimmt wurden (z.B. einzelne Intervalle); hingegen resultieren Experimente mit analytischen Materialien, d.h. mit natürlichen Reizmustern (z.B. Musikstücke), bei weitem häufiger in monoton fallenden

[1] Eine ausführliche Beschreibung und Kritik bis dahin durchgeführter Studien zum Zusammenhang zwischen Reizkomplexität und dem Ausdruck von Wohlgefallen, Freude und Präferenz bietet Raab, der sich dabei jedoch nicht allein auf musikalische Urteile beschränkt (1981, S. 276f.).

Beziehungen zwischen Komplexität und ästhetischem Gefallen (Niketta, 1993, S. 332). Neben der Verwendung unterschiedlicher Materialien als Stimuli unterscheiden sich die Studien ebenfalls darin, ob das Ausmaß an Komplexität subjektiv, d.h. durch die Versuchspersonen selber, oder objektiv, also vom Versuchsleiter anhand von Expertengesprächen, bestimmt wird. Ein direkter Vergleich der Ergebnisse fällt daher schwer. Darüber hinaus ist zu bemerken, dass bei der Interpretation der Ergebnisse in einer Vielzahl der Studien recht großzügig vorgegangen wird und lineare Zusammenhänge zwischen Komplexität und Urteil als ‚Hinweise' auf eine kurvilineare Beziehung interpretiert werden, in der nur entweder der niedrigste oder der höchste Grad an Komplexität berücksichtigt wurde und daher nur der lineare Teil der Funktion zu sehen ist (Raab, 1981, S. 276f.). Eine Falsifikation der Hypothesen wird auf diese Weise vermieden.

Neuartigkeit

Auch zwischen der kollativen Variablen ‚Neuartigkeit' und dem Urteil des Publikums postuliert Berlyne (1971, 1974b) eine umgekehrt U-förmige Beziehung, da zu einem bestimmten Zeitpunkt eine Sättigung erreicht ist.[1] Dies geht ebenfalls auf das 'arousal-Konzept' zurück, nach dem „people prefer moderate, rather than extreme levels of arousal […] and that this is related to the amount of *information* that they receive from different stimuli" (Hargreaves, 1986, S. 116). Auch für diese Hypothese findet Berlyne empirische Belege, die – vergleichbar mit denen der kollativen Variablen ‚Komplexität' (siehe oben) – von anderen Forschern sowohl bestätigt als auch widerlegt wurden (Hargreaves, 1986, S. 110ff.).

Zusammenhang Komplexität – Neuartigkeit

Abschließend zu bemerken zu den „New Experimental Aesthetics" ist, dass die kollativen Variablen ‚Komplexität' und ‚Neuartigkeit' nicht unabhängig nebeneinander stehen (Berlyne, 1974a, S. 67). Hargreaves (1986) stellt ihren Zusammenhang sehr deutlich dar, indem er speziell auf die vom Zuhörer empfundene, subjektive Komplexität eingeht und diese als „function both of its objective complexity, and of the familiarity of the listener with respect to that piece" (Hargreaves, 1986, S. 116) darstellt.

[1] Diese Annahme war jedoch schon damals nicht neu: Schon McDougall (1908) und Piaget (1936) hatten vermutet, dass die Effekte neuartiger Reize ihren höchsten Wert nicht bei maximaler Ausprägung, sondern durch einen mittleren Grad erreichen (Berlyne, 1974, S. 42).

2.2.3.1.3 Der Zuhörer

Ein zweiter Forschungsstrang der Rezeptionsforschung beschäftigt sich mit den Auswirkungen subjektiver Variablen auf die Rezeption von Musik. Im Folgenden steht speziell die emotionale Reaktion eines Zuhörers im Fokus, da in der vorliegenden Arbeit ein Einfluss der emotionalen Wirkung einer Aufführung auf das Gesamturteil des Publikums über einen Abend in der Oper (‚Gesamterlebnis Opernaufführung') vermutet wird.

Zwei bedeutsame Forschungsstränge beschäftigen sich mit der emotionalen Reaktion auf Musikstücke: Neben einem Zweig, der Musik als *Ausdruck* emotionaler Zustände untersucht (Dollase et al., 1986; Lehmann, 1994; Sloboda & O'Neill, 2001; Bersch-Burauel, 2004), widmet sich ein zweiter – deutlich kleinerer – Forschungsstrang der Musik als *Auslöser* von Emotionen. Letzterer erscheint im Kontext der vorliegenden Untersuchung weit mehr von Interesse.

Hülshoff definiert Emotionen als „körperlich-seelische Reaktionen, durch die ein Umweltereignis aufgenommen, verarbeitet, klassifiziert und interpretiert wird, wobei eine Bewertung stattfindet" (Hülshoff, 1999, S. 14). Obwohl es sich um eine allgemeine, d.h. nicht speziell auf Musik bezogene Definition handelt, erscheint sie auch in diesem Kontext sinnvoll. Die Reaktion des Publikums auf ein „Umweltereignis" (hier: Opernaufführung) soll erklärt werden, wobei speziell dessen „Bewertung" (hier: sein Gesamturteil über einen Abend in der Oper) interessiert. Einen ähnlichen Ansatz wählen Sloboda und O'Neill (2001), die Emotionen als Reaktion auf Musik jedoch nicht als „abstract entities such as ‚anger' or ‚elation', but rather as actual moments of emotional feelings and displays in particular situations within a particular culture" sehen (Sloboda & O'Neill, 2001, S. 415).

Nach wie vor bedeutend in diesem Zusammenhang sind die Untersuchungen Maslows, der Versuchspersonen auf der Suche nach *peak experiences* ihre „happiest moments, ecstatic moments, moments of rapture" (Maslow, 1968, S. 71) beschreiben lässt und in Musik (insbes. klassischer Musik) einen häufigen Auslöser findet (Maslow, 1976, S. 169f., vgl. ebenfalls Laski, 1961; Panzarella, 1980). Als Vertreter häufig zitierter Untersuchungen jüngeren Datums ist Csikszentmihalyi (1991) zu nennen, dessen *flow*-Konzept große Ähnlichkeiten mit Maslows *peak experiences* aufweist: Auch Csikszentmihalyi geht es um Erfahrungen „so enjoyable that people will do it […] for the sheer sake of doing it" (Csikszentmihalyi, 1991, S. 4). Als Kennzeichen von *flow* sieht er dabei u.a. die Verschmelzung von Handlung und

Bewusstsein, das Ausblenden des Zeitfaktors sowie ein autotelisches Erleben. Diese können durch eine Vielzahl von Aktivitäten ausgelöst werden, wozu neben Klettern, Tanzen und Musizieren eben auch das Hören eines Musikstückes gehört (Csikszentmihalyi, 1991, S. 68). Wie jedoch lassen sich von Musik ausgelöste Emotionen exakter beschreiben und klassifizieren?

Im Rahmen einer Studie, die Gabrielsson (2001) mit mehr als 300 Personen durchführt, bitte er sie, „the strongest, most intense experience of music that you have ever had“ (Gabrielsson, 2001, S. 434) zu beschreiben. Anschließend teilt er die von den Befragten geschilderten Emotionen in vier Gruppen ein: intensive, positive, negative sowie gemischte bzw. konfligierende Emotionen (2001, S. 435f.). Da jedoch explizit nach *strong experiences of music* gefragt wurde, scheint der Erfolg einer exakten Übertragung auf die vorliegende Untersuchung fraglich.

Eine andere Vorgehensweise wählt Pike (1972), der Laien nach dem Hören verschiedener Musikstücke um Beschreibung ihrer emotionalen Reaktion bittet. Daraus extrahiert er fünf Dimensionen der empfundenen Emotionen: *feeling of pleasure, feeling of oneness with the music, feeling of movement, perception of stabile moods* sowie *perception of spontaneous and transient emotional states* (Pike, 1972, S. 264). Nachdem die drei erstgenannten Dimensionen keiner weiteren Erklärung bedürfen, wird auf die verbleibenden zwei genauer eingegangen: Unter *perception of stabile moods* versteht Pike die generelle Wirkung der Musik und fasst darunter anhaltende Gefühle wie Freude, Trauer oder Sehnsucht nach Vergangenem. Im Gegensatz dazu steht die *perception of spontaneous and transient emotional states*. Diese sind „not indicative of any specific emotion, but are characterized primarily by more general and spontaneous inner feelings than by reasoned or objective calculations [...], sometimes succeeding each other very rapidly“ (Pike, 1972, S. 265f.). Als am häufigsten auftretend werden die Dimensionen *feeling of pleasure*, *feeling of oneness with the music* und *perception of stabile moods* identifiziert (Pike, 1972, S. 264).

Speziell mit der physischen Reaktion auf Musik befassen sich Goldstein (1980), Sloboda (1991, 1999) und Panksepp (1995), wobei sich letzterer speziell dem durch Musik ausgelösten Schauern bzw. Frösteln widmet. Weitere in der Literatur häufig erwähnte Formen physischer Reaktion auf Musik sind Tränen, Zittern, Kloß im Hals, Gänsehaut, veränderte Atmung,

Druck auf der Brust sowie ein beschleunigter Herzschlag (Gabrielsson, 2001, S. 441). Einen ausgezeichneten Überblick bieten Scherer und Zentner (2001, S. 374ff.).[1]

2.2.3.1.4 Die Situation

Neben subjektiven Variablen und dem Werk selbst bestimmen auch situative Bedingungen die Rezeption von Musik. Obwohl es vom überwiegenden Teil der Studien zum Stimulus (Abschnitt 2.2.3.1.2) sowie zu den subjektiven Variablen (Abschnitt 2.2.3.1.3) außer Acht gelassen wird, ist festzuhalten, dass Urteile über eine Aufführung immer im Kontext anderer Menschen gefällt werden – also „im Bewusstsein dessen, wie diese urteilen und wie sie gegebenenfalls unser Urteil bewerten könnten“ (Behne, 1997, S. 1006). Gerade in uneindeutigen Wahrnehmungssituationen muss von dem Einfluss der zuvor abgegebenen Urteile anderer Personen ausgegangen werden. Eine beispielhafte Studie führen Furman und Duke (1988) mit Orchestermusik durch: Nachdem in die Versuchsanordnung Eingeweihte, d.h. Scheinversuchspersonen, ihr Urteil über die Aufführung öffentlich abgegeben hatten, wurden die Zuhörer um ihr Urteil gebeten. Deutliche Suggestionseffekte zeigen sich, die mit zunehmender musikalischer Vorbildung der Befragten schwächer werden (Furmann & Duke, 1988, S. 225). Ebenfalls häufig untersucht wurden sog. ‚Prestige-Effekte’, d.h. Auswirkungen wertender Aussagen von (fiktiven) Experten. Auch in jenen Studien werden signifikante und mäßig bis starke Effekte gezeigt, die zum Großteil nicht vom Ausmaß der musikalischen Vorbildung der Teilnehmer beeinflusst sind (de la Motte-Haber, 1973; Kreitler & Kreitler, 1980; Wilson, 1985; Behne, 1987). Ein Überblick über die ältere Literatur findet sich bei Hargreaves (1986, S. 194ff.), der zusammenfassend bemerkt, dass „[a]ll but one of the studies […] found social influences to have a significant effect on aesthetic judgements“ (Hargreaves, 1986, S. 198).

2.2.3.2 Erklärungen individueller Urteile über einen Abend im Sprechtheater

Nach der Analyse der Literatur zur Erklärung individueller Urteile über einen Abend im Konzert (Abschnitt 2.2.3.1) wird es im Folgenden um die Rezeption von Aufführungen im Sprechtheater gehen. Welche Determinanten bestimmen hier das Publikumsurteil? Unterscheiden sich diese von jenen bei der Musikrezeption? Dazu wird zunächst Eversmann (2004) vorgestellt, der vier Dimensionen der Interaktion zwischen Theaterstück und Zuschauer identifiziert, woraus sich Kriterien ableiten lassen, die das Urteil eines Zuschauers

[1] Zu beachten ist, dass emotionale Reaktionen auf Musik, egal ob psychischer oder physischer Form, in ihrer Art, ihrem Ausmaß und ihrer Richtung in Abhängigkeit von mehreren Faktoren sehr unterschiedlich ausfallen. Zu den wichtigsten Faktoren gehören Eigenschaften des Stimulus selbst, Hörgewohnheiten des Befragten sowie dessen kultureller und situativer Hintergrund (Harrer, 1993, S. 596f.; Gabrielsson, 2001, S. 442ff.).

über eine Theateraufführung bestimmen (Abschnitt 2.2.3.2.1). Da in der vorliegenden Arbeit jedoch das Publikumsurteil über den gesamten Abend im Mittelpunkt steht, schließt sich – analog zur Erklärung individueller Urteile über einen Abend im Konzert – die Präsentation von Literatur an, die sich mit Emotionen bzw. Identifikation im Sprechtheater (Abschnitt 2.2.3.2.2) sowie mit den situativen Bedingungen einer Theateraufführung (Abschnitt 2.2.3.2.3) beschäftigt, da diese in der vorliegenden Arbeit ebenfalls als Determinanten des Gesamturteils des Publikums über den Abend gesehen werden.

2.2.3.2.1 Der Stimulus ‚Aufführung im Sprechtheater'

Eversmann (2004) geht der Frage nach dem Gesamterlebnis einer Theateraufführung nach und wendet dazu eine Arbeit von Csikszentmihalyi und Robinson (1990), die sich mit der Struktur[1] ästhetischer Erfahrungen[2] in der Kunst allgemein beschäftigt, auf das Theater an (Eversmann, 2004, S. 148ff.).

Als erste Dimension sieht Eversmann die *perceptual dimension*, die die reine Auffassung bzw. Wahrnehmung eines Theaterstücks umfasst, d.h. die Erfahrung der Aufführung selber, noch ohne eine Interpretation oder Bedeutungszuweisung. Eine zweite Dimension stellt die *cognitive dimension* beim Publikum dar, d.h. dessen Vorwissen und Kenntnisse über das Stück, das Thema auf der Bühne oder den Regisseur. Die intellektuelle Herausforderung besteht darin, der Handlung zu folgen sowie die Zeichen und Symbole auf der Bühne zu interpretieren. Ebenfalls in diese Dimension fällt laut Eversmann ein Vergleich der Erwartungen des Publikums mit der Realität der Aufführung (vgl. Abschnitt 2.2.3.1.1), wobei die Erwartungen in hohem Maße von den jeweiligen Fachkenntnissen des Zuschauers abhängen. Ebenso bedeutsam innerhalb dieser Dimension ist das ‚Sich Wiedererkennen' des Zuschauers in Charakteren auf der Bühne, welches damit „a key element in the reception process" darstellt (Eversmann, 2004, S. 153). Als dritte Dimension wird die *emotional dimension* beschrieben, die der Autor wiederum unterteilt: An erster Stelle werden Emotionen beim Zuschauer erzeugt, die in Verbindung mit dem (fiktionalen) Inhalt der Aufführung stehen; es handelt sich dabei um ähnliche Emotionen wie die, die die Charaktere auf der Bühne in ihren jeweiligen Situationen empfinden, sowie um solche, die der Zuschauer einem Charakter entgegenbrächte, sollte er diesem im wahren Leben begegnen. Weitere Emotionen

[1] Die Struktur *(structure)* ästhetischer Erfahrungen, bei der große Ähnlichkeiten zwischen Individuen bestehen, steht dabei im Gegensatz zu deren Inhalt *(content)*, der sich von Zuschauer zu Zuschauer deutlich unterscheidet.
[2] Ästhetische Erfahrungen stellen laut Csikszentmihalyi und Robinson (1990) eine spezielle Form des *flows* dar, die in Abschnitt 2.2.3.1.3 näher erläutert wird.

entstehen in Verbindung mit dem Theaterbesuch selber: Diese umfassen laut Eversmann das Interesse an der Vorstellung, das Gefühl, alles um sich herum zu vergessen *(flow)*, sowie Bewunderung und Freude. Auch physische Reaktionen können darin enthalten sein, die „most of the time seem to be absent from the accounts of aesthetic experiences in other art forms" (Eversmann, 2004, S. 156).[1] Als vierte Dimension der Wahrnehmung einer Theateraufführung gilt die *communicative dimension*, die die drei erstgenannten umschließt und die Interaktion zwischen Zuschauer und Aufführung beinhaltet. Offensichtlich an dieser Stelle ist die Kommunikation zwischen dem Publikum und den Akteuren, die jedoch nicht die einzige ist: So kann das Publikum ebenfalls mit dem Regisseur, dem Verfasser des Stückes, dessen Kultur und Epoche sowie untereinander interagieren. Zusätzlich wird in einigen von Eversmann geführten Interviews von einer inneren Kommunikation berichtet, während der die Zuschauer „seem to become actively engaged in an internal debate about the issues raised by the performance" (Eversmann, 2004, S. 158).

Auch ohne die explizite Erwähnung durch den Autor selbst lassen sich aus seiner Anwendung der von Csikszentmihalyi und Robinson (1990) herausgearbeiteten vier Dimensionen auf das Sprechtheater Kriterien ableiten, die das Urteil eines Zuschauers über eine Theateraufführung bestimmen. So wird einfließen, ob seine erste Reaktion dem Thema, dem Hauptdarsteller oder dem Bühnenbild gegenüber positiv oder negativ war *(perceptual dimension)* und ob er in der Lage war, der Handlung des Stückes zu folgen – ohne angestrengt oder im Gegenteil intellektuell unterfordert zu sein *(cognitive dimension)*. Ebenfalls wird in sein Gesamturteil einfließen, ob er während der Aufführung Emotionen empfand, die denen der Akteure auf der Bühne entsprachen, inwieweit den Alltag hinter sich lassen konnte *(emotional dimension)* und ob die Aufführung in der Lage war, ihn zum Nachdenken anzuregen *(communicative dimension)*.

2.2.3.2.2 Die emotionale Wirkung einer Theateraufführung auf den Zuschauer

Nachdem die emotionale Wirkung einer Aufführung auf den Zuschauer als eine Determinante seines individuellen Urteils identifiziert wurde, werden im Folgenden weitere Arbeiten vorgestellt, die sich speziell mit Emotionen im Sprechtheater beschäftigen, da in der vorliegenden Arbeit von einem bedeutenden Einfluss der Emotionen auf das Gesamturteils des Publikums ausgegangen wird.

[1] Dieser Aussage muss widersprochen werden, da auch bei der Rezeption von Musik von – unterschiedlich starken – physischen Reaktionen berichtet wurde (vgl. Abschnitt 2.2.3.1.3).

Konijn (1999) identifiziert anhand einer Studie während eines Theaterfestivals in den Niederlanden (n = 757) drei Dimensionen von Emotionen: Identifikation, Empathie[1] und *task emotions* (Konijn, 1999, S. 173 ff.). Ihr Verständnis von Identifikation geht dabei auf Aristoteles' Konzept der Katharsis (übersetzt 1968) zurück und umfasst Emotionen, die der Zuschauer parallel zu denen der Charaktere auf der Bühne in ihren jeweiligen Situationen empfindet (vgl. Abschnitt 2.2.3.2.1). Unklar bleibt ihre Auffassung von Empathie, die sich nur wenig von der Identifikation abzuheben scheint (Konijn, 1999, S. 173).[2] Einen wesentlich größeren Einfluss auf das Gesamterlebnis schreibt Konijn den *task emotions* zu, „that result from the task of spectating, such as ‚I should concentrate better on the plot, but what a well-performed Hedda!' The understanding of the game that is played by the actors on stage […] stirs task emotions in spectators" (Konijn, 1999, S. 173). Diese können sowohl negativ als auch positiv sein, wobei unter negative Emotionen u.a. Verwirrung, Verärgerung und Langeweile fallen; zu den positiven gehören Konzentration, Herausforderung, Spannung und Bewunderung. Parallelen zu Berlynes „New Experimental Aesthetics" (u.a. 1960, 71, 74a, 74b, vgl. Abschnitt 2.2.3.1.2) fallen ins Auge.

Drei Auslöser von Emotionen benennt Konijn: Neben der Fiktion auf der Bühne und den dazugehörigen Gegenständen können auch persönliche Erinnerungen und Assoziationen Emotionen erzeugen, die sich demnach nicht zwangsläufig direkt auf Ereignisse auf der Bühne beziehen müssen (Konijn, 1999, S. 174).[3] Bei der Auswertung der empirischen Ergebnisse zeigt sich, dass eine Identifikation mit den Charakteren auf der Bühne kaum stattfindet, Empathie und positive *task emotions* hingegen in hohem Ausmaß auftreten (Konijn, 1999, S. 189).

Ebenfalls mit dem Prozess der Identifikation im Sprechtheater setzen sich Schoenmakers und Tulloch (2004) auseinander, die dabei den Einfluss der spezifischen Aufführung auf das Ausmaß der Identifikation deutlich machen. Drei entscheidende Faktoren werden herausgearbeitet: Für eine starke Identifikation des Zuschauers mit Charakteren auf der Bühne muss erstens ein Gefühl der Ähnlichkeit zwischen ihm und dem entsprechenden Charakter

[1] Als englischsprachige Wendung des Originalbegriffs ‚Einfühlung' setzte sich der Begriff ‚empathy' bzw. ‚Empathie' in der Literatur durch (Kreitler & Kreitler, 1980, S. 252).
[2] Einen ausführlichen Überblick über Gemeinsamkeiten und Unterschiede der Begriffe ‚Identifikation' und ‚Empathie' bietet Zillmann (1991, S. 135f.).
[3] Im Gegensatz zu Eversmann (2004), der persönliche Erinnerungen und Assoziationen zur *cognitive dimension* zählt (vgl. Abschnitt 2.2.3.2.1), schreibt Konijn (1999) diese den Emotionen (genauer: den Auslösern von Emotionen) zu.

entstehen, das sich auf Alter und Geschlecht sowie auf Persönlichkeit und Werte bezieht.[1] Zweitens muss der Charakter eine gewisse Attraktivität ausstrahlen, so dass beim Betrachter der Wunsch entsteht, diesem ähnlich zu sein. Als dritte Bedingung einer starken Identifikation gelten formale Aspekte der Aufführung. So werden sich Zuschauer bspw. nur schwerlich mit Charakteren identifizieren, die sehr kleine Rollen spielen und/oder über die kaum Informationen erhältlich sind (Schoenmakers & Tulloch, 2004, S. 19f.).

Einen anderen Weg wählt Zillmann (1994), den ebenfalls die Prozesse interessieren, die (teilweise sogar intensive) Gefühlsbeteiligungen der Zuschauer an Geschehnissen und Schicksalen auf der Bühne vermitteln, wenngleich sie nur fiktiv sind: „How can so-called rational beings fall prey to the actors' personas and respond to them as they were real persons in their immediate environment […]?" (Zillmann, 2004, S. 34). Er zweifelt am Nutzen des Identifikations-Konzeptes zur Erklärung emotionaler Reaktionen im Theater (siehe oben), nachdem er zeigt, dass sich das Publikum mit Bühnenfiguren nicht identifiziert, sondern sich diesen gegenüber verhält „in much the same way they respond to friends and foe in their actual, immediate social environment" (Zillmann, 1994, S. 36) (vgl. Abschnitt 2.2.3.2.1). Von weit größerem Nutzen zur Erklärung von Emotionen im Theater erscheint ihm die Empathie, die er folgendermaßen definiert:

> „Empathy is defined as any experience that is a response (a) to information about circumstances presumed to cause acute emotions in another individual and/or (b) to the bodily, facial, paralinguistic, and linguistic expression of emotional experiences by another individual and/or (c) to another individual's actions that are presumed to be precipitated by acute emotional experiences, this response being (d) associated with an appreciable increase in excitation and (e) construed by respondents as feeling with or feeling for another individual." (Zillmann, 1994, S. 40)

Nach einer Diskussion verschiedener Aspekte der Empathie, auf deren Erläuterung an dieser Stelle aus Platzgründen verzichtet wird, schildert Zillmann ihren Einfluss auf die Rezeption einer Theateraufführung: So müssen die Charaktere für eine emotionale Reaktion der Zuschauer so gestaltet sein, dass diese entweder als Freund oder Feind interpretiert werden

[1] Dem widerspricht Eversmann (2004), nach dem sich das Publikum in den meisten Fällen mit dem Hauptdarsteller identifiziert und „outward characteristics, such as age, gender or race, only play a minor role here." (Eversmann, 2004, S. 153f.)

können. Hier gilt laut Zillmann ‚je stärker, desto besser' insofern, als dass „the more the respondents are emotionally touched and taken in by dramatic events, the more likely it is that they will appraise, in retrospect, the drama experience as positive and enlightening" (Zillmann, 1994, S. 48). Ein klarer Einfluss der Emotionen auf das Publikumsurteil im Sprechtheater wird demnach unterstellt.

2.2.3.2.3 Die Situation

Auch situative Bedingungen gelten in der Literatur als Determinanten des Gesamturteils des Publikums über einen Abend im Theater. Diese können in den Einfluss anderer Menschen sowie den der Rahmenbedingungen unterteilt werden.

So werden Urteile auch im Sprechtheater im Kontext anderer Menschen gefällt und gerade in „nicht eindeutigen Wahrnehmungssituationen [können] zuvor abgegebene Urteile anderer das eigene Urteil beeinflussen" (Behne, 1997, S. 1006, vgl. ebenfalls Wilson, 1985, S. 65ff.). Dazu gehören u.a. in den Medien veröffentlichte Theaterkritiken, die insbesondere die von Csikszentmihalyi und Robinson (1990) betitelte *cognitive dimension* der Wahrnehmung (vgl. Abschnitt 2.2.3.2.1) insofern beeinflussen, als dass vor der Aufführung Erwartungen beim Zuschauer geweckt werden und dieser während der Aufführung bzw. danach seine eigenen Gedanken und Interpretationen mit denen des Kritikers bzw. der Kritiker abgleicht (Eversmann, 2004, S. 168). Ähnliches gilt für die Reputationen der Schauspieler, des Regisseurs, des Bühnenbildners oder Dramaturgen, die die Erwartungen im Voraus ebenso beeinflussen können. Zusätzlich hat auch die Länge des Applauses vor der Pause bzw. insbesondere nach Ende der Vorstellung einen Einfluss in dem Sinne, dass „spectators are contaminated by emotions from others and/or that feelings are intensified and reinforced by the reaction of their neighbours" (Eversmann, 2004, S. 171; vgl. ebenfalls Wilson, 1985, S. 65f.).

Einen erheblichen Einfluss auf das Urteil über einen Abend im Sprechtheater wird auch den Rahmenbedingungen zugesprochen, zu denen das Theatergebäude, seine Lage, die Atmosphäre im Theater und der Komfort der Sitze genauso gehören wie die Informationen, die die Zuschauer von Seiten des Theaters erhalten (Programmheft, etc.) (Eversmann, 2004, S. 165; Kreitler & Kreitler, 1980, S. 250). Drei Funktionen schreibt Eversmann diesen zu: „During the performance the (spatial) conditions can help to focus the concentration, to further communication with the performers and to establish a sense of unity" (Eversmann,

2004, S. 166). Bei Betrachtungen des Gesamterlebnisses eines Theaterabends fallen weitere Faktoren ins Gewicht: So sind auch die Wahl des Stückes, die Infrastruktur (öffentliche Verkehrsmittel, Parkmöglichkeiten, etc.), der Empfang des Publikums beim Betreten des Gebäudes oder das Vorhandensein von Diskussionsmöglichkeiten nach der Vorstellung nicht ohne Einfluss (Eversmann, 2004, S. 166). Anders als bei den Subjekt-Variablen der Zuschauer können die Verantwortlichen des Theaters hier also einen sehr guten Zustand herzustellen versuchen und sind in der Lage, das Publikumsurteil positiv zu beeinflussen – ohne (übermäßig) hohe Kosten und ohne Gefährdung des künstlerischen Anspruchs.[1]

2.3 Determinanten des Gesamturteils des Publikums über einen Abend in der Oper (‚Gesamterlebnis Opernabend')

Aus Mangel an Forschung zu den Determinanten des Gesamturteils des Publikums über einen Abend in der Oper (‚Gesamterlebnis Opernabend') wurden in den vorangehenden Abschnitten Literatur zur Erklärung individueller Urteile über einen Abend im Konzert (vgl. Abschnitt 2.2.3.1) sowie über einen Abend im Sprechtheater (vgl. Abschnitt 2.2.3.2) vorgestellt. Zur Klärung der Untersuchungsfragen 3 (Determinanten des Gesamturteils) und 4 (Unterschiede im Gesamturteil zwischen Laien und Experten) werden nun Determinanten des Gesamturteils des Publikums über einen Abend in der Oper daraus abgeleitet. Es wird von drei Kriteriengruppen ausgegangen: die wahrgenommene künstlerische Qualität (Abschnitt 2.3.1), die individuelle Reaktion des Zuschauers (Abschnitt 2.3.2) sowie die situativen Rahmenbedingungen (Abschnitt 2.3.3). Abschnitt 2.3.4 widmet sich den Beziehungen der Determinanten untereinander und gibt diese grafisch wieder.

2.3.1 Die wahrgenommene künstlerische Qualität

Als Kriterium mit erheblichem Einfluss auf das Gesamturteil des Publikums über einen Abend in der Oper wird die künstlerische Qualität der zu bewertenden Aufführung gesehen. Zu ihrer Erhebung kann auf das Modell Boerners (2002) zurückgegriffen werden (vgl. Abschnitt 2.2.1). Da die Arbeiten von Moser (2005), Renz (2006) und Boerner et al. (in Druck) erkennen lassen, dass das analytische Modell das Qualitätsurteil des Publikums nicht vollständig abzubilden in der Lage ist, werden in der vorliegenden Arbeit einige Konstrukte modifiziert werden (vgl. Abschnitt 3.1.2).

[1] zum konstanten Zielkonflikt der öffentlichen Theater zwischen künstlerischem Anspruch, ökonomischen Notwendigkeiten und Wünschen des Publikums vgl. Abschnitt 1.1

2.3.2 Individuelle Reaktion des Zuschauers

Neben dem Einfluss der wahrgenommenen Qualität der Aufführung wird ebenfalls die individuelle Reaktion des Zuschauers als Determinante seines Gesamturteils über einen Abend in der Oper gesehen. Dazu gehören die Emotionen des Zuschauers, sein Eindruck von der Komplexität und Neuartigkeit der Aufführung sowie seine Identifikation mit den Charakteren auf der Bühne.

Emotionale Wirkung

Sowohl bei Konzertaufführungen als auch bei Vorstellungen im Sprechtheater zeigt sich bei der Literaturanalyse ein erheblicher Einfluss der emotionalen Wirkung auf das Publikumsurteil, so dass von diesem auch bei Opernaufführungen auszugehen ist (vgl. Behr, 1983, S. 164ff.). Für die vorliegende Arbeit wird in Anlehnung an Pike (1972) und Konijn (1999) von sechs Dimensionen der Emotionen ausgegangen, die eine Opernaufführung in den Zuschauern auslöst. Diese sind ‚Empathie', ‚persönliche Assoziationen', ‚Genussfaktor', ‚Ich-Beteiligung', ‚vorübergehende Gefühlsregungen' und ein ‚Gefühl des Bewegt-Seins'. Unter ‚Empathie' wird dabei in Anlehnung an Zillmann (1994) ein Mitfühlen mit den Darstellern verstanden. Ein beispielhaftes Item im Fragebogen lautet: ‚Während der Aufführung fieberte ich mit dem Schicksal der Darsteller mit'. Im Gegensatz dazu geht es bei den ‚persönlichen Assoziationen' um Emotionen, die im Zuschauer ausgelöst werden, weil dieser an Situationen seines eigenen Lebens erinnert wird. Ebenfalls mit Einfluss auf das Publikumsurteil wird der ‚Genussfaktor' gesehen, bei dem die Stimmung des Zuschauers im Mittelpunkt steht. Ein Item lautet: ‚Während der Aufführung war ich froh und entspannt'. Um das *flow*-Konzept von Csikszentmihalyi (1991) geht es bei der ‚Ich-Beteiligung' und es wird erhoben, ob der Zuschauer während der Aufführung in der Lage war, alles um ihn herum zu vergessen. ‚Vorübergehende Gefühlsregungen' beschreiben kurzfristige Reaktionen auf das Bühnengeschehen (Pike, 1972), worunter auch physische Reaktionen wie Herzklopfen und feuchte Hände fallen. Im Gegensatz dazu steht das ‚Gefühl des Bewegt-Seins', das über kurzfristige Gefühlsregungen hinausgeht und eventuell auch nach der Vorstellung anhält.

Kollative Variablen

Wie bereits erwähnt wird auch den Variablen ‚Komplexität' und ‚Neuartigkeit' der Aufführung ein Einfluss auf das Gesamturteil des Publikums über einen Abend in der Oper eingeräumt. Ein Zusammenhang zwischen dem Ausmaß der empfundenen Komplexität bzw.

Neuartigkeit und den Fachkenntnissen sowie der Erfahrung des einzelnen Zuschauers wird vermutet.

Identifikation des Publikums

Anders als Konijn (1999), die die Identifikation des Publikums mit einzelnen Charakteren zu den durch die Aufführung ausgelösten Emotionen zählt (vgl. Abschnitt 2.2.3.2.2), wird diese in der vorliegenden Arbeit – analog zu Eversmann (2004) – als eigenes Konstrukt erfasst. Nur auf diese Weise kann ermittelt werden, ob sich das Publikum mit den Akteuren auf der Bühne identifiziert und in welchem Ausmaß dies in sein Gesamturteil einfließt.

2.3.3 Situative Rahmenbedingungen

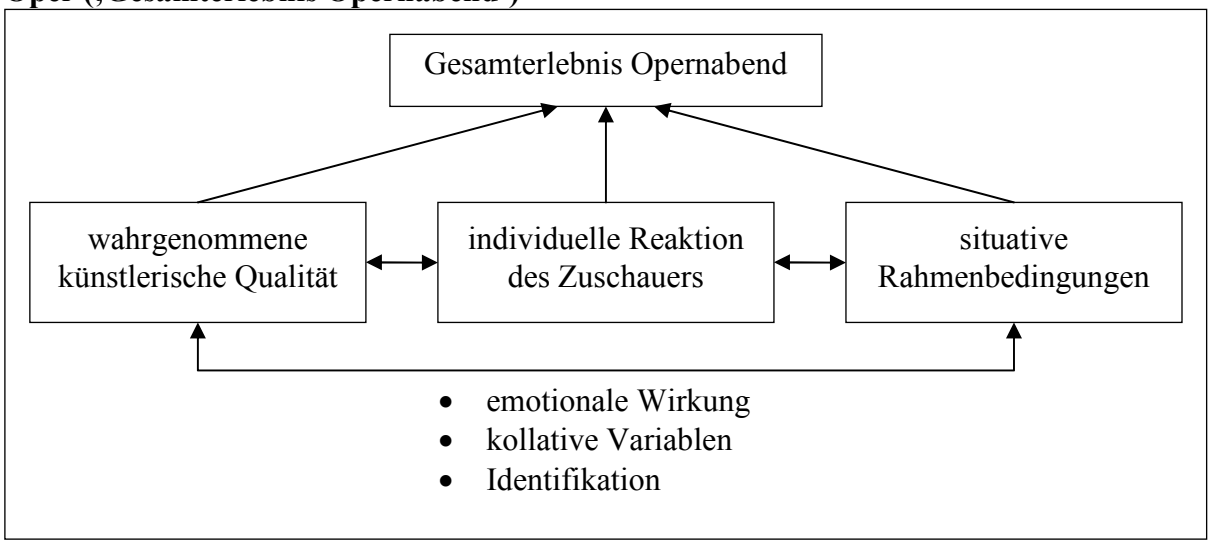
Als eine weitere Determinante des Gesamturteils des Publikums über einen Abend in der Oper werden Rahmenbedingungen identifiziert, die weder in der Aufführung selber noch in den Zuschauern liegen. Hierunter fallen in der vorliegenden Arbeit die Akustik des Saales, der Komfort des Sitzes, die Sicht auf die Bühne, störende Ablenkungen während der Aufführung, die Temperatur im Zuschauerraum, der Service des Hauses, Räumlichkeiten neben dem Aufführungssaal selber (Foyer, etc.) und die Länge der Aufführung.

Ebenfalls zu den situativen Rahmenbedingungen würde der Einfluss der Urteile anderer Personen auf das eigene Urteil gehören (vgl. Abschnitte 2.2.3.1.4 und 2.2.3.2.3), dessen Erhebung im Rahmen der vorliegenden Untersuchung jedoch nicht möglich ist.

2.3.4 Das ‚Gesamterlebnis Opernabend'

In der vorliegenden Arbeit wird angenommen, dass die Determinanten nicht unabhängig voneinander sind, sondern sich gegenseitig beeinflussen. So könnte bspw. die Wahrnehmung der musikalischen Leistung des Orchesters beeinträchtigt sein, sollte die Akustik im Aufführungssaal nicht dem Standard entsprechen. Ebenfalls könnten ständige Geräusche aus dem Publikum einen Zuschauer daran hindern, sich auf die Aufführung zu konzentrieren und ihm so die Möglichkeit nehmen, die Gefühle der Charaktere auf der Bühne nachzuempfinden. Auch die Identifikation mit Charakteren könnte beeinträchtigt sein, sollte bspw. das Bühnenbild die Aufführung in Bezug auf die Handlung und Darsteller nicht unterstützen. Abbildung 3 gibt sämtliche in dieser Arbeit angenommenen Determinanten des Gesamturteils des Publikums über einen Abend in der Oper sowie ihre Beziehungen untereinander grafisch wieder.

Abbildung 3: Determinanten des Gesamturteils des Publikums über einen Abend in der Oper (‚Gesamterlebnis Opernabend')



3. Instrument und Methode

3.1 Entwicklung des Fragebogens zur Erfassung der Determinanten des Gesamturteils des Publikums über einen Abend in der Oper (‚Gesamterlebnis Opernaufführung')

Wie erwähnt dient ein Teil der vorliegenden Arbeit der Validierung des von Moser (2005) und Renz (2006) entwickelten Fragebogens zur Erfassung des Publikumsurteils über die künstlerische Qualität im Musiktheater (vgl. Abschnitt 2.2.2). Dabei ist das Ziel, nur noch Konstrukte abzufragen, die vom Publikum für die Bewertung eines Abends in der Oper als relevant betrachtet werden, sowie ausschließlich solche Items zu verwenden, die ein durchschnittliches Opernpublikum kompetent zu beantworten in der Lage ist. Dazu werden Items, bei denen die Befragten Schwierigkeiten mit der Beantwortung hatten, aus dem Fragebogen entfernt, sowie Konstrukte des analytischen Modell Boerners (2002) überarbeitet (Abschnitt 3.1.2). Darüber hinaus wird in der vorliegenden Arbeit nicht mehr allein die künstlerische Qualität eine Aufführung betrachtet, sondern es wird versucht, die Determinanten des Gesamturteils des Publikums über einen Abend in der Oper (‚Gesamterlebnis Opernabend') zu ermitteln. Daher werden in den Fragebogen weitere Konstrukte aufgenommen, wie in Abschnitt 3.1.1 erläutert.[1]

3.1.1 Aufbau und Antwortformat des Fragebogens

Der Fragebogen setzt sich aus den vier Bestandteilen ‚Gesamteindruck', ‚Bewertung der Einzelaspekte', ‚Bewertung des Zusammenspiels' und ‚Abschließende Fragen' sowie einer letzten Frage zum Zeitpunkt der Beantwortung zusammen. Bezüglich der Aspekte der Bewertung der künstlerischen Qualität einer Opernaufführung wird dabei die Struktur des analytischen Modells von Boerner (2002), an der sich auch der Fragebogen von Renz (2006) orientiert, grundsätzlich beibehalten. Hinzu kommen Aspekte, welche sich auf die Determinanten des Gesamturteils des Publikums über einen Abend in der Oper (‚Gesamterlebnis Opernabend') beziehen (vgl. Abschnitt 2.3). Abbildung 4 gibt die Struktur des Fragebogens grafisch wieder.

[1] Der Fragebogen der vorliegenden Studie befindet sich in Anhang A. Zur Vereinfachung eines Vergleiches mit dem Fragebogen der Studie von Renz (2006), der in Anhang B zu finden ist, sind die neu konstruierten bzw. veränderten Items grau hinterlegt. Ebenfalls in Anhang A befinden sich die Kennziffern der Items zum besseren Verständnis.

Abbildung 4: Aufbau des Fragebogens

A. Gesamteindruck		
1. Gesamterlebnis Opernaufführung (GE)		
2. künstlerische Qualität der Aufführung (kQ)		
3. Inszenierung (Komplexität, Vertrautheit, Art der Inszenierung, Identifikation)		
4. Emotionen (Empathie, persönliche Assoziation, Genussfaktor, Ich-Beteiligung, vorübergehende Gefühlsregung, Gefühl des Bewegt-Seins)		
B. Bewertung der Einzelaspekte		
I. Musikalische Leistung des Ensembles	II. Szenische Leistung des Ensembles	
1. Orchester, Zusammenspiel des Orchesters, Dirigent (MO, fit3_O, MD) 2. Solist (MSolist) 3. Chor (MChor)	1. Solist (SSolist) 2. Chor (SChor) 3. Ausstattung (SAusstattung)	
C. Bewertung des Zusammenspiels		
1. Zusammenspiel innerhalb des Chores (fit3_Chor)		
2. Zusammenspiel der Solisten untereinander (fit3_Solisten)		
3. Zusammenspiel von Orchester, Solisten und Chor (fit2_Musik)		
4. Stimmigkeit der szenischen Aspekte (fit2_Szene, fit3_Ausstattung)		
5. Stimmigkeit zwischen Musik und Szene (fit1)		
D. Abschließende Fragen		
Rahmenbedingungen Akustik, Sitz, Sicht, Ablenkungen, Temperatur, Service, Räumlichkeiten, Atmosphäre, Länge der Aufführung	**Persönliche Angaben** Alter, Geschlecht, Schulabschluss, Opernerfahrung, Besuchshäufigkeit, Abonnement, Dauer des Abonnements, Instrument, Kenntnisse des Stücks, Stimmung vor/ während/ nach der Aufführung	**Einschätzung der Wichtigkeit** Orchester, Solisten, Chor, Dirigent, Regisseur, Bühnenbildner, Techniker, Beleuchter
E. Letzte Frage: Zeitpunkt der Beantwortung		

Zur Erfassung der Determinanten des Gesamturteils des Publikums über einen Abend in der Oper (‚Gesamterlebnis Opernabend') teilt sich der Abschnitt ‚Gesamteindruck' in die Unterpunkte ‚Gesamterlebnis Opernabend' und ‚künstlerische Qualität der Aufführung'. Ebenfalls unter den Abschnitt ‚Gesamterlebnis' fallen die Unterpunkte ‚Inszenierung' und ‚Emotionen'. Der Unterpunkt ‚Inszenierung' setzt sich dabei aus den Skalen ‚Komplexität', ‚Vertrautheit', ‚Gesamtkonzept' und ‚Identifikation' zusammen[1]; sechs Bestandteile hat der Unterpunkt ‚Emotionen', deren theoretischer Hintergrund in den Abschnitten 2.2.3.1.3 und

[1] Die Überschrift ‚Inszenierung' wurde hierbei weniger aus inhaltlichen Überlegungen als aus Gründen der Übersichtlichkeit und Verständlichkeit gewählt.

2.2.3.2.2 erläutert wird: ‚Empathie', ‚persönliche Assoziation', ‚Genussfaktor', ‚Ich-Beteiligung', ‚vorübergehende Gefühlsregungen' und ‚Gefühl des Bewegt-Seins'.

Es folgt der Abschnitt ‚Bewertung der Einzelaspekte', der sich entsprechend dem analytischen Modell Boerners (2002) aus den Teilen ‚musikalische Leistung des Ensembles' und ‚szenische Leistung des Ensembles' zusammensetzt. Als einzige Akteure sind dabei das Orchester sowie der Dirigent ausschließlich bei der musikalischen Leistung zu finden, was sich durch ihre – im Normalfall – fehlende szenische Funktion erklären lässt. Im Unterschied zum analytischen Modell sind das Orchester (MO), das Zusammenspiel des Orchesters (fit3_O) sowie der Dirigent (MD) im Fragebogen der vorliegenden Studie zusammengefasst, wie in Abschnitt 3.1.2 näher erläutert werden wird.

Um die Teilnehmer bei der Beantwortung des Fragebogens nicht zu überfordern, beschränken sich die Items der Konstrukte ‚musikalische Leistung des Solisten' (MSo) und ‚szenische Leistung des Solisten' (SSo) auf *einen* Solisten, der jeweils vorgegeben wurde.[1]

Der Fragebogen setzt sich fort mit der ‚Bewertung des Zusammenspiels', das sich auf alle drei Ebenen des analytischen Modells von Boerner (2002) bezieht. Auf der dritten Ebene befinden sich dabei das ‚Zusammenspiel innerhalb des Chores' (fit3_C) sowie das ‚Zusammenspiel der Solisten untereinander' (fit3_S). Ebenfalls an dieser Stelle steht im analytischen Modell sowie dementsprechend im Fragebogen von Renz (2006) das ‚Zusammenspiel innerhalb des Orchesters' (fit3_O), welches im Fragebogen der vorliegenden Studie wie oben erwähnt mit der Bewertung des Einzelaspektes ‚Orchester' zusammengelegt wird. Ursachen und Hintergründe sind in Abschnitt 3.1.2 zu finden. Ebenfalls unter der ‚Bewertung des Zusammenspiels' befinden sich das ‚Zusammenspiel von Orchester, Solisten und Chor' (fit2_Musik) und die ‚Stimmigkeit der szenischen Aspekte' (fit2_Szene, fit3_Ausstattung), die die zweite Ebene des analytischen Modells von Boerner (2002) bilden.[2] Auch die ‚Stimmigkeit zwischen Musik und Szene' (fit1), welche die erste Ebene des analytischen Modells darstellt, findet sich in diesem Teil des Fragebogens.

[1] Ebenfalls ein Solist wurde für das Unterkonstrukt ‚Identifikation' unter ‚A. Gesamteindruck, 3. Inszenierung' ausgewählt. Dabei wurde neben dem Solisten jedoch ebenfalls die Solistin erwähnt, um die Beantwortung des Items unabhängig von der Geschlechterrolle zu machen. So lautet eines der Items ‚Das Verhalten des Solisten bzw. der Solistin ist für mich vorbildlich'.

[2] Besonderheiten der ‚Stimmigkeit der szenischen Aspekte' werden ebenfalls in Abschnitt 3.1.2 erläutert.

Der abschließende Teil des Fragebogens besteht aus Fragen zu den Rahmenbedingungen der Opernaufführung, Fragen nach persönlichen Angaben der Teilnehmer sowie zur Einschätzung der Wichtigkeit der verschiedenen Akteure. Obwohl sie nicht explizit zur Erfassung der Wahrnehmung der künstlerischen Qualität benötigt wird, dient letztgenannte Frage speziell der Erfassung der Wichtigkeit des Dirigenten, worauf in Abschnitt 3.1.2 näher eingegangen werden wird.[1]

Zur Beantwortung der Fragen stand eine Likert-Skala zur Verfügung, die wegen ihrer einfachen Konstruktion und Anwendung die am häufigsten verwendete Skalierungsmethode in den Sozialwissenschaften darstellt (Schnell et al., 2005, S. 191). Die Befragten konnten zwischen fünf Antwortmöglichkeiten wählen, die vom Fragebogen der Studien von Renz (2006) und Boerner et al. (in Druck) übernommen wurden: „trifft voll und ganz zu“, „trifft eher zu“, „teils, teils“, „trifft eher nicht zu“ und „trifft überhaupt nicht zu“. Nach der Klassifikation von Stevens (1946) wurde somit das Skalenniveau der Intervallskala erreicht.[2] Darüber hinaus standen bei der Mehrzahl der Items die Antworten „habe ich nicht beachtet“ sowie „kann ich nicht beurteilen“ zur Verfügung.[3] Auf diese Weise sollten weitere Erkenntnisse darüber gewonnen werden, auf welche Kriterien sich das Urteil des Publikums über die künstlerische Qualität einer Aufführung tatsächlich stützt. Die Antwortmöglichkeit „habe ich nicht beachtet“ bestand für den Fall, dass bestimmte Items bei der Bewertung der künstlerischen Qualität bzw. des Gesamterlebnisses von Einzelnen nicht wahrgenommen wurden. Wenn Items jedoch wegen fehlenden Wissens nicht beantwortet werden konnten, konnten die Befragten die Antwortmöglichkeit „kann ich nicht beurteilen“ ankreuzen. Beide Antwortkategorien sind als gültig zu interpretieren und unterscheiden sich somit von ‚echten’ fehlenden Werten.[4] Bei speziellen statistischen Berechnungen (z.B. Reliabilitätsanalysen) stellen sie jedoch ebenfalls fehlende Werte dar.

[1] Nach welchen Rahmenbedingungen bzw. demographischen Angaben genau gefragt wurde, lässt sich Abbildung 4 entnehmen. Der genaue Wortlaut der Fragen findet sich in Anhang A.

[2] Zusätzlich zur Unterscheidungsmöglichkeit sowie zur Rangordnungsmöglichkeit erfordern Intervallskalen, dass die Intervalle stets die gleiche Größe besitzen, d.h. die zugeordneten Zahlen müssen immer dieselbe Differenz aufweisen.

[3] Auch diese wurden von den Untersuchungen von Renz (2006) und Boerner et al. (in Druck) übernommen, bei denen in nicht unerheblichem Ausmaß von den Antwortmöglichkeiten Gebrauch gemacht wurde.

[4] ‚Echte’ fehlende Werte, d.h. Items, bei denen keine der gegebenen Antwortmöglichkeiten angekreuzt wurde, kommen in der vorliegenden Untersuchung kaum vor und sind daher zu vernachlässigen. Zusätzlich wurden die Fragebögen aus der Analyse ausgeschlossen, bei denen mehr als 20% der Items nicht beantwortet wurden (Roth, 1994). Daher werden im Folgenden die Antworten „habe ich nicht beachtet“ bzw. „kann ich nicht beurteilen“ als fehlende Werte bezeichnet.

Entsprechend den Empfehlungen der Literatur, nicht ausschließlich Items zu verwenden, bei denen die Zustimmung eine positive Einstellung widerspiegelt (Schnell et al., 2005, S. 187), wurden auch Items konstruiert, bei denen die Ablehnung eine positive Einstellung ausdrückt.[1] Damit soll die Zustimmung zu einer Frage ohne jeglichen Bezug zu deren Inhalt offen gelegt werden, was als ‚Zustimmungstendenz' oder ‚Akquieszenz' bezeichnet wird (Schnell et al., 2005, S. 354f.).

3.1.2 Modifikation der Skalen und Items

Wie im vorhergehenden Abschnitt erwähnt, wurden neben der Neukonstruktion der abhängigen Variablen ‚Gesamterlebnis Opernabend' auch Skalen und Items der Variablen ‚künstlerische Qualität einer Aufführung in der Oper' verändert, wobei insbesondere die Analysen der prozentualen Häufigkeiten nicht beantworteter Items in der Studie von Renz (2006) sowie der von Boerner et al. (in Druck) und ihre Faktoren- sowie Reliabilitätsanalysen als Hilfestellung dienten. Wie bereits erläutert, zielten diese Modifikationen auf die Erhöhung der Validität des Fragebogens.

Neuentwicklung der Skala 'Gesamterlebnis Opernabend'

Da in der vorliegenden Arbeit die Determinanten des Gesamturteils des Publikums über einen Abend in der Oper (‚Gesamterlebnis Opernabend') bestimmt werden sollen (Untersuchungsfragen 3 und 4) und nicht mehr ausschließlich die künstlerische Qualität der Aufführung betrachtet wird (Untersuchungsfragen 1 und 2), musste der entsprechende Teil des Fragebogens ergänzt werden (siehe Anhang A, S. b, Items mGE1 – vNG2). Neben der Neukonstruktion von Items für die latente Variable ‚Gesamterlebnis Opernabend' wurden die Items zum Gesamturteil über die künstlerische Qualität ebenfalls neu formuliert (Ausnahmen: urGU1, vNG2[2]). Grund hierfür war neben einer klaren Abgrenzung beider Skalen die Tatsache, dass die Skala ‚Gesamturteil' in der statistischen Auswertung der Studie von Renz (2006) bei der Faktorenanalyse ohne Vorgabe von Faktoren in drei Faktoren zerfiel, was eine zu starke Ausdifferenzierung vermuten lässt. Entfernt wurden darüber hinaus sämtliche Items, die die emotionale Wirkung der Aufführung abfragten, da diese im Fragebogen der vorliegenden Untersuchung eine eigene Skala bildet (vgl. Abschnitt 2.3.2).

[1] Beispiele sind das Item ‚Der Abend hat mich sehr enttäuscht' sowie ‚Die Inszenierung war mir insgesamt zu neuartig und modern'.
[2] Die Formulierungen der Items können Anhang A entnommen werden.

Neuentwicklung der Skala 'Inszenierung'

Auch die Skala ‚Inszenierung', die sich bei Renz (2006) sowie bei Boerner et al. (in Druck) aus der Bewertung der Inszenierung selbst sowie aus kollativen Variablen zusammensetzt, wurde nach der Analyse ihrer Ergebnisse komplett überarbeitet (siehe Anhang A, S. b – c, Items rKA1_k – mII3_i). Vier Teilaspekte bilden nun die Skala ‚Inszenierung': ‚Komplexität', ‚Vertrautheit', ‚Gesamtkonzept' und ‚Identifikation'. Alle vier Teilaspekte wurden mit jeweils drei Items operationalisiert, die bis auf drei Ausnahmen komplett von Renz (2006) übernommen werden konnten.[1]

Neuentwicklung der Skala ‚Emotionen'

Komplett neu erarbeitet wurde die Skala ‚Emotionen', die aus den sechs theoretisch hergeleiteten und in Abschnitt 2.3.2 beschriebenen Teilaspekten ‚Empathie', ‚persönliche Assoziationen', ‚Genussfaktor', ‚Ich-Beteiligung', ‚vorübergehende Gefühlsregungen' und ‚Gefühl des Bewegt-Seins' besteht (siehe Anhang A, S. c – d, Items mEE1 – mrEM3).

Neuentwicklung der Skala ‚Musik_Orchester'

Im Gegensatz zum analytischen Modell Boerners (2002), das die drei Konstrukte ‚musikalische Leistung des Orchesters' (MO), ‚Zusammenspiel innerhalb des Orchesters' (fit3_O) sowie ‚Leistung des Dirigenten' (MD) trennt, wurden diese in der vorliegenden Arbeit zu einer Skala zusammengefasst (siehe Anhang A, S. e, Items mMO13 – MD13). Dies wird begründet durch die hohe Anzahl nicht beantworteter Items dieser Konstrukte in den Studien von Renz (2006) und Boerner et al. (in Druck), wo sich zeigte, dass das Publikum zu einer Differenzierung in diesem Ausmaß nicht in der Lage ist.

Die verhältnismäßig hohe Anzahl nicht beantworteter Items der Konstrukte ‚musikalische Leistung des Orchesters' (MO) und ‚Zusammenspiel innerhalb des Orchesters' (fit3_O) machte eine spätere Skalenbildung bei Renz (2006) unmöglich, vorausgesetzt eine Skala sollte aus mindestens drei Items bestehen. Sie nennt als möglichen Grund, dass zur Beurteilung des Orchesters weniger die emotionale Einschätzung als vielmehr faktisches Wissen nötig sei. Dies könnte eine Erklärung dafür sein, weshalb sich das Publikum aufgrund des höheren Abstraktgehaltes durchaus in der Lage sah, das Konstrukt ‚Stimmigkeit innerhalb

[1] Neu konstruiert wurden die Items ‚Das Verhalten des Solisten bzw. der Solistin ist für mich vorbildlich', ‚Ich konnte mich mit der Rolle des Solisten bzw. der Solistin sehr gut identifizieren' und ‚Der Solist bzw. die Solistin stehen für Werte, die mir persönlich sehr wichtig sind', die das Unterkonstrukt ‚Identifikation' operationalisieren.

der musikalischen Dimension' (fit2_Musik) zu beurteilen (Renz, 2006, S. 29), wo das Orchester ebenfalls eine wichtige Rolle spielt.[1] Auffallend ist der hohe Anteil der Antwortkategorie „habe ich nicht beachtet" bei der Beurteilung der musikalischen Leistung des Orchesters speziell durch die Laien. Die Erwartung von Renz (2006), dass aufgrund fehlenden musikalischen Wissens und mangelnder Hörerfahrung die Laien das Orchester *nicht beurteilen* können, wurde nicht erfüllt; vielmehr scheint das Orchester für die Laien keine große Rolle für ihre Beurteilung zu spielen, da sich „für einen musikalischen Laien vermutlich fast jedes professionelle Orchester ‚gut' anhört" (Renz, 2006, S. 34).[2]

Ähnliches gilt für das Konstrukt ‚Leistung des Dirigenten' (MD), bei dem der Anteil nicht beantworteter Items sowohl bei der Befragung in Konstanz (Renz, 2006, S. 134) als auch bei der in Köln (Boerner et al., in Druck) noch höher als beim Orchesters lag. Die Mehrzahl der betreffenden Items wurde dabei nicht beantwortet, weil das Publikum aufgrund fehlenden Wissens und Erfahrung die Leistung des Dirigenten *nicht beurteilen* konnte. Dies widerspricht der Auffassung Jungheinrichs (1986), der die häufige Nicht-Beachtung des Dirigenten in der Oper mit der Vielzahl anderer Akteure erklärt, mit denen der Dirigent – anders als bei Konzertaufführungen – um die Aufmerksamkeit des Publikums konkurriert (Jungheinrich, 1986, S. 168). Darüber hinaus ist denkbar, dass das Publikum den Dirigenten nicht beurteilen konnte, weil dieser während der Aufführung kaum zu sehen war, „was bei schlundartig tiefer Orchesterversenkung und dann, wenn der Dirigent kein Hüne ist" (Jungheinrich, 1986, S. 168), nicht verwunderlich ist. Doch ebenfalls bei Opernhäusern, in denen der Dirigent für das Publikum sichtbar ist, werden von der Autorin ähnliche Ergebnisse vermutet, da die Leistung des Dirigenten für die überwiegende Mehrheit der Zuschauer wohl nicht von der des Orchesters, des Chores sowie der Solisten zu trennen ist.[3]

[1] Items des Konstruktes 'fit2_Musik' sind bspw. 'Orchester, Chor und Solisten waren in Bezug auf die Lautstärke sehr gut aufeinander abgestimmt' und 'Die Balance zwischen dem Orchester und den Sängern war sehr gut gelungen, so dass weder das Orchester noch die Sänger ‚zugedeckt' wurden'.

[2] Um dieser Frage gesondert nachzugehen, wurde für den Fragebogen der vorliegenden Untersuchung eine Frage konstruiert, die explizit nach der Wichtigkeit des Orchesters für die künstlerische Qualität einer Opernaufführung fragt. Ergebnisse liefert Tabelle 6 (Abschnitt 4.1.3).

[3] Obwohl ein großer Teil des Publikums zu einer separaten Bewertung der Leistung des Dirigenten nicht in der Lage ist, wird dieser als wichtig für die künstlerische Qualität einer Opernaufführung allgemein, d.h. unabhängig von der bewerteten Aufführung, gesehen (vgl. Items mW1 – mW8 im Fragebogen, S. m, Anhang A). So steht der Dirigent an dritter Stelle nach den Solisten und dem Orchester (siehe Tabelle 6, Abschnitt 4.1.3). Deutliche Unterschiede ergeben sich bei einer getrennten Betrachtung der Laien und Experten: Nimmt der Dirigent in der Teilstichprobe der Laien den vierten Platz ein, schätzen ihn die Experten als wichtigsten Akteur für die künstlerische Qualität einer Opernaufführung ein (siehe Anhang G, Tabellen 30 und 31).

Ein weiterer Grund für die Zusammenlegung der Konstrukte ‚musikalische Leistung des Orchesters' (MO), ‚Zusammenspiel innerhalb des Orchesters' (fit3_O) und ‚Leistung des Dirigenten' (MD) liegt in den Zusammenhängen der Skalen von Moser (2005) begründet, die er anhand einer Korrelationsmatrix zeigt. Dabei korreliert die Skala ‚musikalische Leistung des Orchesters' sowohl stark mit der Skala ‚Zusammenspiel innerhalb des Orchesters' ($r = .68$, $p < .01$) als auch mit der Skala ‚Leistung des Dirigenten' ($r = .64$, $p < .01$). Zwischen den Skalen ‚Zusammenspiel innerhalb des Orchesters' und ‚musikalischen Leistung des Dirigenten' besteht eine mittlere Korrelation ($r = .42$, $p < .01$), so dass eine Zusammenlegung insgesamt gerechtfertigt erscheint.[1]

Skala ‚Szene_Ausstattung'

Hingegen wurde an der Skala ‚Szene_Ausstattung', die nach Boerners analytischem Modell (2002) auf zwei Ebenen zu finden ist, wenig verändert: Auf der zweiten Ebene befindet sich die ‚Stimmigkeit der szenischen Aspekte' (fit2_Szene), wohingegen die Ausstattung auf der dritten Modellebene sowohl hinsichtlich der Einzelaspekte (z.B. Bühnenbild, SA) als auch in Hinblick auf das Zusammenspiel der Ausstattungseffekte (fit3_A) ausdifferenziert werden kann (Boerner, 2002, siehe Abbildung 1). Im Fragebogen der vorliegenden Untersuchung wurde dabei jedoch wie bei Renz (2006) und Boerner et al. (in Druck) auf die gesonderte Erhebung der Stimmigkeitsaspekte der dritten Ebene (fit3_Ausstattung) verzichtet. Die entsprechenden Items sind im Fragebogen zur besseren Übersicht unter ‚C. Bewertung des Zusammenspiels – Stimmigkeit der szenischen Aspekte' (fit2_Szene) zu finden (siehe Anhang A, S. 9, Items fit3A1 – vfit2S_3). In Übereinstimmung mit dem analytischen Modell Boerners (2002) wird die Skala ‚Szene_Ausstattung' (SA) wie schon bei Renz (2006) getrennt erhoben. Dieses umfasst im Gegensatz zur Version von Renz (2006) und Boerner et al. (in Druck) im Fragebogen der vorliegenden Untersuchung nur noch die Aspekte, die auch unabhängig von der Stimmigkeit einen ästhetischen Wert haben (siehe Anhang A, S. g, Items vSA13 – mSA17).[2]

[1] Ebenfalls bestätigt wird dieses Vorgehen durch die daran anschließende Faktorenanalyse, die Moser (2005) anhand der Skalen durchführt. Orchester und Dirigent lassen sich dabei auch faktorenanalytisch nicht unterscheiden (Moser, 2005, S. 54).

[2] Aus diesem Grund entfiel bspw. das Item ‚Der Einsatz der Bühnentechnik hat mir gut gefallen', dessen Beantwortung einer so erheblichen Zahl von Rezipienten schwer fiel, dass bei Boerner et al. (in Druck) keine Skala ‚Szene_Ausstattung' gebildet werden konnte.

Weitere Änderungen

Neben den genannten, zum Teil grundlegenden Änderungen der Skalen wurden weitere Änderungen vorgenommen, die dem Ziel der Verständlichkeit dienten. So wurde bspw. der Begriff ‚Produktion' trotz unterschiedlicher Bedeutung durch ‚Aufführung' ersetzt. Ebenfalls gestrichen wurde der Begriff ‚Werktreue', der mehrdeutig und in der Literatur als Kriterium umstritten ist (u.a. Boerner, 2002, S. 112).

3.2 Datenerhebung

Zur Durchführung der Befragung wurde eine Opernaufführung am Anhaltischen Theater Dessau gewählt, das mit 1.096 Plätzen im Großen Haus, 150 Plätzen im Foyer und etwa 80 Plätzen im Studio sowie einer Gesamtbühnenfläche von 1450 m² das größte Theater Sachsen-Anhalts ist. Es wurde in der zweiten Hälfte des 18. Jahrhunderts gegründet und ist heute ein Mehrspartentheater mit Musiktheater (Oper, Operette, Musical), Schauspiel, Tanztheater und der Anhaltischen Philharmonie. Deutschlandweit bekannt ist der psychologische Realismus des Musiktheaters des Generalintendanten Johannes Felsenstein, welches den singenden Menschendarsteller ins Zentrum rückt und alle inszenatorischen Vorgänge auf der Basis des von Komponist und Librettist geschaffenen Werkes erarbeitet.

Für die Durchführung der Befragung wurde die Oper „Die Zauberflöte" von Wolfgang Amadeus Mozart gewählt, die bspw. in der Spielzeit 2004/05 541 Mal an deutschen Bühnen aufgeführt und dabei von mehr als 320.000 Besuchern gesehen wurde (Deutscher Bühnenverein, 2006, S. 70). Mit 45 Inszenierungen allein in der genannten Spielzeit ist „Die Zauberflöte" damit die am häufigsten inszenierte Oper Deutschlands.[1] Nicht allein wegen ihrer Bekanntheit jedoch wurde „Die Zauberflöte" für die vorliegende Untersuchung ausgewählt, sondern vielmehr aufgrund der konventionellen Art, mit der sie in Dessau inszeniert wurde. Dies liegt darin begründet, dass die in Boerners analytischem Modell (2002) theoretisch angenommenen Konstrukte wohl eher in konventionellen Inszenierungen zu beobachten sind, was erst ermöglicht, die Struktur des Gesamturteils auf Grundlage des hierarchischen Modells vollständig zu erfassen.

Wie in Abschnitt 3.1.1 erläutert, beziehen sich die Items der Konstrukte ‚musikalische Leistung des Solisten' (MSo) und ‚szenische Leistung des Solisten' (SSo) im Fragebogen auf

[1] Den zweiten Platz belegt mit großem Abstand die Oper „Hänsel und Gretel" von Engelbert Humperdinck, die in der Spielzeit 2004/05 25 Mal inszeniert wurde und in 262 Aufführungen gut 150.000 Besucher anzog (Deutscher Bühnenverein, 2006, S. 70).

einen Solisten. Für die Durchführung der vorliegenden Studie wurde die Rolle des Papageno gewählt, der – im Gegensatz zur Königin der Nacht oder zu Pamino – nahezu konstant auf der Bühne präsent ist. Darüber hinaus ist die Rolle des Papageno in der ausgewählten Produktion eher unkonventionell inszeniert (u.a. Kostüm, Frisur), wodurch ein hoher Erinnerungswert garantiert sein sollte.

Vor Beginn der gewählten Aufführung, die am 22. Juni 2007 stattfand, wurde das Publikum durch Plakate im Foyer und Durchsagen durch den Inspizienten auf die Befragung aufmerksam gemacht. Darüber hinaus bekam jeder Zuschauer beim Betreten des Zuschauerraums einen Text überreicht, der ebenfalls über das Stattfinden der Befragung und ihre Initiatoren informierte.[1] Um die Ergebnisse durch eine erhöhte Aufmerksamkeit bzw. verschärfte Wahrnehmung des Publikums nicht zu verzerren, wurde das genaue Ziel der Studie entsprechend den Empfehlungen der Literatur (Butzer-Strothmann et al., 2001) an dieser Stelle nicht genannt. Die Fragebögen selber wurden mit Hilfe des Abendpersonals des Theaters in der Pause verteilt, wobei versucht wurde, jeden Zuschauer persönlich anzusprechen und gezielt um die Teilnahme an der Befragung zu werben.[2]

Zum Ausfüllen des Fragebogens bestanden drei Möglichkeiten: in der Pause, nach der Aufführung im Theater oder zu Hause. Da zu vermuten war, dass potentielle Teilnehmer nicht bereit wären, nach Ende der Aufführung noch länger im Theater zu verweilen, sollte auf diese Weise die Stichprobe vergrößert werden. Um die Befragten nicht durch anfallende Portokosten zu belasten, wurden bei Bedarf zusätzlich adressierte Freiumschläge verteilt. Zur Steigerung der Attraktivität einer Teilnahme wurden darüber hinaus Freikarten für das Anhaltische Theater Dessau unter den Teilnehmern verlost.

Insgesamt wurde somit sowohl bei der Konstruktion des Fragebogens als auch bei der Durchführung der Befragung nach der von Dillman (1978, 2007) entwickelten ‚Total-Design-Method'[3] vorgegangen. Diese analysiert die potentiellen Schwierigkeiten, die sich aus der Konstruktion des Fragebogens sowie aus der Erhebungssituation ergeben können, und betrachtet die Teilnahme bzw. Nicht-Teilnahme an einer Befragung als Kosten-Nutzen-Kalkül der potentiell Befragten. Dillmans Methode besteht nun darin, den Befragten den

[1] Der Wortlaut der Ankündigung ist in Anhang I zu finden.
[2] Auf diese Weise sollte sowohl die Stichprobengröße erhöht als auch vermieden werden, dass sympathisch bzw. zugänglich wirkende Personen überproportional oft ausgewählt wurden (Reuband, 2005, S. 124).
[3] Diese wurde inzwischen von Dillman in ‚Tailored-Design-Method' umbenannt (Dillman, 2007).

möglichen Nutzen einer Teilnahme zu vermitteln und gleichzeitig die entstehenden Kosten (u.a. Zeitaufwand, finanzielle Aufwendungen) so gering wie möglich zu halten (Dillman, 2007).[1]

3.3 Vorgehensweise der Datenauswertung

Zur Auswertung der erhobenen Daten, die mit dem Statistikprogramm SPSS 15.0 durchgeführt wird, werden zunächst einfaktorielle Varianzanalysen gerechnet, um zu klären, ob Unterschiede in den Bewertungen abhängig vom Ausfüll-Zeitpunkt bestehen (vgl. Abschnitt 3.2). Sollte dies der Fall sein, könnte die Stichprobe nicht als Ganzes betrachtet werden und separate Datenanalysen wären erforderlich. Es folgt eine deskriptive Auswertung der Häufigkeit des Auftretens der Antwortmöglichkeiten „habe ich nicht beachtet“ bzw. „kann ich nicht beurteilen“. Ziel dabei sind Erkenntnisse über das Ausmaß der Übereinstimmung zwischen der Struktur der empirisch wahrgenommenen Aufführungsqualität und dem hierarchischen Modell (Boerner, 2002). Zur Überprüfung der Güte des modifizierten Fragebogens schließen sich Reliabilitäts- und Validitätsanalysen an. Zu letzteren gehören sowohl die Betrachtung der Interkorrelation der gebildeten Skalen als auch Faktorenanalysen, mit denen nach der Überprüfung der Güte des Fragebogens zusätzlich Erkenntnisse zu den Untersuchungsfragen 1a (musikalische und szenische Aspekte einer Opernaufführung) und 2a (Unterschiede zwischen Laien und Experten in Bezug auf Differenziertheit und Gewichtung der Teilaspekte) gewonnen werden. Zur Beantwortung der Untersuchungsfrage 1b (Gewichtung der Teilaspekte) folgen Pfadanalysen und Strukturgleichungsmodelle. Anschließend werden Unterschiede zwischen den Qualitätsurteilen von Laien und Experten untersucht (Untersuchungsfrage 2), wozu separate Pfadanalysen, T-Tests sowie Intraklassenkorrelationen verwendet werden.

Zur Beantwortung der Untersuchungsfrage 3, was Gegenstand des Gesamturteils des Publikums über einen Abend in der Oper ist, werden wiederum Pfadanalysen durchgeführt. Anschließend werden die Bewertungen von Laien und Experten getrennt analysiert (Untersuchungsfrage 4). Mit T-Tests wird versucht zu klären, ob Teilaspekte des Gesamturteils unterschiedlich bewertet werden (Untersuchungsfrage 4a). Anschließend wird die Höhe der Übereinstimmung der Urteile innerhalb der Laien- und der Expertengruppe anhand von Intraklassenkorrelationen überprüft (Untersuchungsfrage 4b). Kapitel 5 schließt

[1] Ursprünglich entwickelt wurde die ‚Total-Design-Methode’ für postalische Befragungen, wobei ein großer Teil der Ausführungen ebenfalls für andere Formen der Befragung äußerst hilfreich sind.

mit Pfadanalysen zu den Unterschieden zwischen Laien und Experten in Bezug auf die Gewichtung der Teilaspekte (Untersuchungsfrage 4c).

4. Vorbereitende Analysen

4.1 Deskriptive Statistiken

Nach der Beschreibung der Stichprobe der vorliegenden Untersuchung (Abschnitt 4.1.1) werden einfaktorielle Varianzanalysen durchgeführt, um eventuelle Unterschiede je nach Ausfüll-Zeitpunkt des Fragebogens festzustellen (Abschnitt 4.1.2). Es folgt eine deskriptive Auswertung der Daten, wobei Erkenntnisse zum Ausmaß der Übereinstimmung zwischen der Struktur der empirisch wahrgenommenen Aufführungsqualität und dem hierarchischen Modell Boerners (2002) erwartet werden (Abschnitt 4.1.3). Anschließend wird die Bildung des Indexes ‚Expertentum' erläutert (Abschnitt 4.1.4), der in Kapitel 5 zur Klärung der Untersuchungsfragen 2 und 4 Verwendung finden wird.

4.1.1 Beschreibung der Stichprobe

Im Rahmen der vorliegenden Untersuchung wurden 120 Personen zwischen 18 und 75 Jahren befragt. Bei 560 Besuchern der Aufführung der Oper „Die Zauberflöte" am entsprechenden Abend ergibt sich eine Rücklaufquote von gut 32%, geht man davon aus, dass an zwei Drittel aller Besucher ein Fragebogen mit der Bitte um eine Teilnahme an der Befragung verteilt wurde. Da in vergleichbaren Untersuchungen, d.h. Befragungen sowohl mit der Möglichkeit einer Rückgabe vor Ort als auch einer kostenfreien Zusendung, gewöhnlich Werte zwischen 33 und 40% – bezogen auf die Anzahl ausgeteilter Fragebögen – erreicht werden (Reuband, 2005, S. 125), kann die hier erreichte Quote als zufrieden stellend bezeichnet werden. Als möglicher Grund für das Ausbleiben einer höheren Rücklaufquote wird das Verteilen des Fragebogens durch das Abendpersonal des Theaters gesehen, was die Nachdrücklichkeit, mit der die Bedeutung der Befragung vermittelt wurde, womöglich einschränkte (Reuband, 2005, S. 125f.).

Das Durchschnittsalter der Befragten lag bei 49 Jahren (Standardabweichung = 16). Dies ist vergleichbar zu den Studien von Neuhoff (2001) und Reuband (2005), die bei Untersuchungen des Berliner bzw. Kölner Opernpublikums einen Altersdurchschnitt von 50 bzw. 55 Jahren erhielten (Neuhoff, 2001, S. 20; Reuband, 2005, S. 127; vgl. ebenfalls Nolte et al., 2001, S. 11). 58% der Befragten der vorliegenden Untersuchungen waren weiblich, so dass sich auch das Verhältnis von weiblichen und männlichen Besuchern im Rahmen des Üblichen bewegt (Behr, 1983, S. 127; Dollase et al., 1986, S. 37; Nolte et al., 2001, S. 13; Reuband, 2005, S. 130).

4.1.2 Varianzanalysen

Da es bei der Befragung um die Beurteilung der künstlerischen Qualität der gesamten Opernaufführung sowie um das Gesamturteil des Publikums über den Abend in der Oper ging, wäre eigentlich nur eine Befragung nach Ende der Aufführung sinnvoll. Um den Stichprobenumfang zu vergrößern, bestanden für das Publikum – wie erwähnt – jedoch drei Möglichkeiten zum Ausfüllen des Fragebogens: in der Pause, nach der Aufführung im Theater oder zu Hause (vgl. Abschnitt 3.2). Folgende Häufigkeitsverteilung ergab sich (siehe Tabelle 3):

Tabelle 3: Häufigkeitsverteilung der Ausfüll-Zeitpunkte

	Häufigkeit	Prozent	gültige Prozente	kumulierte Prozente
in der Pause	41	34.2	34.2	34.2
nach der Aufführung	52	43.3	43.3	77.5
zu Hause	23	19.2	19.2	96.7
teils in der Pause, teils nach der Aufführung	2	1.7	1.7	98.3
teils in der Pause, teils zu Hause	2	1.7	1.7	100.0
gesamt	120	100.0	100.0	

Unterschiede in der Wahrnehmung je nach Zeitpunkt, zu dem der Fragebogen ausgefüllt wurde, sind nicht auszuschließen. So schreibt Eversmann, dass bspw. der kognitive Prozess, in welchem „subjects concentrate on what might be termed the more theoretical and art-historical aspects of the work of art" (Eversmann, 2004, S. 146), erst nach Ende der Aufführung einsetzt, „when there is time to think over what has been seen, to come to terms with it and to relate the show to the individual's own life" (Eversmann, 2004, S. 153, vgl. Abschnitt 2.2.3.2.1). Unterschiede in der Wahrnehmung sind demnach zu erwarten. Darüber hinaus bestand nur für die Befragten, die den Fragebogen zu Hause ausfüllten, die Möglichkeit einer ausführlichen Kommunikation mit anderen Opernbesuchern über die Aufführung, was deren Urteil ebenfalls beeinflusst haben könnte (Eversmann, 2004, S. 159). Ob tatsächlich Unterschiede zwischen den Gruppen bestehen, muss somit explizit überprüft werden, denn anschließende Datenanalysen müssten jeweils getrennt durchgeführt und interpretiert werden, sollten sich signifikante Unterschiede in der Wahrnehmung der künstlerischen Qualität bzw. des Gesamturteils über den Opernabend ergeben. Eine geeignete Methode stellen einfaktorielle Varianzanalysen dar, welche die Hypothese testen, derzufolge

eine Variable in verschiedenen Teilgruppen der Befragten einen gleich hohen Mittelwert aufweist (Brosius, 2006, S. 495).[1] [2]

Der bspw. für die Skala ‚Stimmigkeit innerhalb der musikalischen Dimension' (fit2_M) bestimmte F-Wert von 1.519 kann sich mit einer Wahrscheinlichkeit von 22.4% auch dann ergeben, wenn alle betrachteten Teilgruppen den gleichen Mittelwert aufweisen.[3] Die Nullhypothese kann folglich nicht zurückgewiesen werden. Ähnlich hohe Signifikanzen ergeben sich für nahezu alle Skalen, so dass nicht von Unterschieden zwischen den Teilgruppen ausgegangen werden muss und bei der Datenanalyse sämtliche Fälle unabhängig vom Ausfüll-Zeitpunkt gemeinsam betrachtet werden können (n = 120).[4] Tabelle 4 zeigt die Ergebnisse einfaktorieller Varianzanalysen beispielhaft für ausgewählte Skalen. Eine vollständige Auflistung ist in Anhang C zu finden.

[1] Dabei wird von der Überlegung ausgegangen, dass sich die Streuung der betrachteten Skalen gedanklich in die Streuung innerhalb der einzelnen Teilgruppen auf der einen und die Streuung zwischen den Teilgruppen auf der anderen Seite aufteilen lässt (zur Bildung der Skalen vgl. Abschnitt 4.2.1). Aus dem Vergleich der Quadratsummen dieser beiden Streuungen schließt die Varianzanalyse auf die Mittelwerte der Grundgesamtheit. Unter Einbezug der jeweiligen Freiheitsgrade wird zu diesem Zweck der F-Wert ermittelt, dessen Zufallsverteilung bekannt ist. In einem weiteren Schritt wird dann die Wahrscheinlichkeit berechnet, mit der F mindestens einen bestimmten Wert annimmt, wenn die Nullhypothese, derzufolge die Mittelwerte der Skalen in allen drei betrachteten Fallgruppen gleich hoch sind, korrekt ist (Signifikanz) (Schnell et al., 2005, S. 457).

[2] Da einfaktorielle Varianzanalysen voraussetzen, dass die Teilgruppen zumindest annähernd gleich groß sind (Bortz, 1999, S. 249f.), werden die vier Fälle ausgeschlossen, in denen der Fragebogen entweder teils in der Pause, teils nach der Aufführung oder teils in der Pause, teils zu Hause ausgefüllt wurde.

[3] zur Vorgehensweise bei der Skalenbildung vgl. Abschnitt 4.2.1

[4] Da einfaktorielle Varianzanalysen, wie oben erwähnt, streng genommen voraussetzen, dass die einzelnen Teilgruppen gleich groß sind, in diesem Fall also die gleiche Anzahl an Personen enthalten, sind die Testergebnisse zunächst vorsichtig zu interpretieren. Aus diesem Grund wurde zur Absicherung eine Zufallsstichprobe gezogen, in der die Teilgruppen gleich groß waren (jeweils n = 23). Da auch hier keine signifikanten Unterschiede zwischen den Mittelwerten der Skalen gefunden wurden, können bei der Datenanalyse tatsächlich sämtliche Fälle unabhängig vom Ausfüll-Zeitpunkt gemeinsam betrachtet werden.

Tabelle 4: Einfaktorielle Varianzanalysen (Faktorvariable: Zeitpunkt des Ausfüllens) – ausgewählte Skalen

		Quadrat-summe	Freiheits-grade	Mittel der Quadrate	F	Signifi-kanz
kQ	zwischen den Gruppen	6.993	2	3.497	3.626	.030
	innerhalb der Gruppen	108.957	113	.964		
	gesamt	115.950	115			
fit2_M	zwischen den Gruppen	1.398	2	.699	1.519	.224
	innerhalb der Gruppen	48.791	106	.460		
	gesamt	50.189	108			
fit2_S	zwischen den Gruppen	2.041	2	1.020	1.160	.318
	innerhalb der Gruppen	91.525	104	.880		
	gesamt	93.566	106			
Orchester	zwischen den Gruppen	.135	2	.067	.118	.889
	innerhalb der Gruppen	48.025	84	.572		
	gesamt	48.160	86			
SChor	zwischen den Gruppen	.891	2	.446	.361	.698
	innerhalb der Gruppen	129.614	105	1.234		
	gesamt	130.505	107			

Anmerkungen: kQ = künstlerische Qualität, fit2_M = Stimmigkeit innerhalb der musikalischen Dimension, fit2_S = Stimmigkeit innerhalb der szenischen Dimension, Orchester = Musikalische Leistung und Zusammenspiel des Orchesters sowie Leistung des Dirigenten, SChor = Szenische Leistung des Chores.

4.1.3 Deskriptive Auswertung

Um Erkenntnisse darüber zu erhalten, welche Aspekte einer Opernaufführung von Zuschauern nicht beachtet werden bzw. nicht beurteilt werden können, soll die Häufigkeit des Auftretens der Antwortmöglichkeiten „habe ich nicht beachtet“ bzw. „kann ich nicht beurteilen“ herangezogen werden. Diese wurden vom Fragebogen, den Renz (2006) in ihrer Untersuchung verwendet hatte, übernommen, um Informationen darüber zu erhalten,

inwieweit das analytische Modell Boerners (2002) mit der Struktur der empirisch wahrgenommenen Qualität einer Opernaufführung übereinstimmt.[1]

Insgesamt nahm die Anzahl der Items, die von einem Teil der Befragten nicht beantwortet wurde, im Vergleich zu den Untersuchungen von Renz (2006) und Boerner et al. (in Druck) erheblich ab. Als Grund wird vermutet, dass eine Vielzahl der Items, bei denen das Publikum in den Studien von Renz (2006) bzw. Boerner et al. (in Druck) Schwierigkeiten mit der Beantwortung hatte, im Zuge der in der vorliegenden Arbeit durchgeführten Modifizierung des Fragebogens nicht mehr aufgenommen bzw. vereinfacht wurde. Erklärtes Ziel dabei war es wie erläutert, den Fragebogen so weiterzuentwickeln, dass er nur noch Dimensionen bzw. Konstrukte enthält, die vom Publikum sowohl beachtet als auch beurteilt werden können. So waren bspw. sechs der zehn Items, die die Leistung des Dirigenten messen sollten, aus diesem Grund aus dem Fragebogen der vorliegenden Studie entfernt worden.[2] Trotzdem wählten bei 24% der Items, die die Antwortmöglichkeiten „habe ich nicht beachtet“ bzw. „kann ich nicht beurteilen“ boten[3], mindestens 5% der Befragten eine dieser Möglichkeiten.

Tabelle 5 zeigt die prozentualen Häufigkeiten nicht beantworteter Items anhand ausgewählter Konstrukte.[4] Zur besseren Übersicht sind die Items fett markiert, die von mindestens 5% der Befragten nicht beantwortet werden konnten. Unterstrichen sind die prozentualen Häufigkeiten, bei denen mindestens 70% der Befragten, die eine der beiden genannten Antwortmöglichkeiten nutzen, die Antwort „kann ich nicht beurteilen“ wählten.

[1] Eine genauere Erläuterung der Antwortkategorien „habe ich nicht beachtet“ bzw. „kann ich nicht beurteilen“ ist in Abschnitt 3.1.1 zu finden.

[2] Als weitere Gründe für die stark zurückgegangene Nutzung der Antwortmöglichkeiten „habe ich nicht beachtet“ bzw. „kann ich nicht beurteilen“ im Vergleich zur Untersuchung von Renz (2006) wird die Tatsache gesehen, dass die Teilnehmer bei der vorliegenden Untersuchung eine ganze Oper ‚live’ gesehen hatten, wohingegen den Befragten der Untersuchung von Renz (2006) Teile von Aufführungen auf Video gezeigt worden waren. Des Weiteren dürften das höhere Durchschnittsalter des Publikums und die damit einhergehende größere Opernerfahrung verantwortlich sein.

[3] Die Konstrukte ‚Gesamteindruck’ und ‚Szene_Ausstattung’ (Ebene 3) sowie die abschließenden Fragen boten die genannten Antwortmöglichkeiten nicht, da vom Vorliegen einer „substantiellen“ Antwort (Schnell et al., 2005, S. 337) ausgegangen werden konnte.

[4] Eine Aufstellung der Häufigkeiten nicht beantworteter Items für sämtliche Items und Konstrukte findet sich in Anhang K.

Tabelle 5: Prozentuale Häufigkeit der nicht beantworteten Items – ausgewählte Konstrukte

Kennziffer des Items	fehlend (in Prozent aller gültigen Antworten)	davon (in % der Fehlenden)	
		„habe ich nicht beachtet“	„kann ich nicht beurteilen“
Musikalische Leistung und Zusammenspiels des Orchesters sowie Leistung des Dirigenten			
mMO13	4.2	0.0	100.0
vMO2	**7.5**	33.3	66.7
vuMO3	1.7	0.0	100.0
vMO5	--	--	--
MO10	--	--	--
vfit3O11	**5.0**	16.7	83.3
fit3O5	**5.9**	0.0	100.0
MD1	**9.2**	9.1	90.9
rMD4	**7.6**	0.0	100.0
vMD7	**6.7**	37.5	62.5
MD13	3.3	25.0	75.0
Musikalische Leistung des Chores			
MC2	**5.0**	33.3	66.7
mMC10	**5.0**	50.0	50.0
MC8	**5.0**	16.7	83.3
MC7	**5.0**	16.7	83.3
Stimmigkeit zwischen Musik und Szene			
vfit1_1	**6.0**	14.3	85.7
rfit1_2	3.5	0.0	100.0
fit1_3	3.4	0.0	100.0
fit1_6	2.6	0.0	100.0
fit1_7	2.6	0.0	100.0
vfit1_8	2.6	0.0	100.0

Anmerkung: Die Formulierungen der entsprechenden Items können Anhang A entnommen werden.

Bei Betrachtung der Häufigkeiten der nicht beantworteten Items fällt auf, dass sich deren Anzahl je nach Konstrukt deutlich unterscheidet. Kaum Probleme hatten die Befragten, Items der Konstrukte ‚musikalische Leistung des Solisten’ (MSo), ‚szenische Leistung des Solisten’ (SSo), ‚szenische Leistung des Chores’ (SC), ‚Zusammenspiel innerhalb des Chores’

(fit3_Chor), ‚Zusammenspiel aller Solisten untereinander' (fit3_Solisten), ‚Zusammenspiel von Orchester, Solisten und Chor' (fit2_Musik) sowie ‚Stimmigkeit des szenischen Aspekte' (fit2_Musik) zu beurteilen. Schwierigkeiten bereitete den Teilnehmern der Befragung hingegen, Indikatoren[1] der Leistung des Orchesters (MO), des Zusammenspiels des Orchesters (fit3_O) sowie der Leistung des Dirigenten zu beantworten (MD; siehe Tabelle 5). Ursächlich dafür könnten fehlendes Wissen und mangelnde Erfahrung sein, die zum Beantworten der Items nach Einschätzung der Befragten benötigt werden.[2] Dies wird bestätigt durch die Tatsache, dass der überwiegende Teil der fehlenden Items *nicht beurteilt* werden konnte und nicht deshalb nicht beantwortet wurde, weil die Befragten das Orchester bzw. den Dirigenten *nicht beachtet* hätten. Ebenfalls verhältnismäßig hohe Anzahlen nicht beantworteter Items weist das – auch auf der dritten Ebene des hierarchischen Modell Boerners (2002) liegende – Konstrukt ‚musikalische Leistung des Chores' (MC) auf, bei dem sämtliche Items von 5% der Befragten nicht beantwortet wurden. Auch hier liegt die Vermutung nahe, dass fehlendes Wissen bzw. mangelnde Erfahrung es den Befragten nach eigener Einschätzung unmöglich machte, die Items zu beantworten.[3]

Nicht begründen lässt sich damit allerdings, warum die wenigsten Teilnehmer Schwierigkeiten bei der Beantwortung der Items zur musikalischen Leistung des Solisten Papageno hatten, wofür vermutlich ebenfalls musikalisches Vorwissen und Erfahrung benötigt wurden.[4] Renz (2006) führt die häufigere Beurteilung der musikalischen Leistung des Solisten auf seine szenische Präsenz auf der Bühne zurück (Renz, 2006, S. 34), die sowohl dem Orchester als auch dem Dirigenten in der Oper üblicherweise sowie dem Chor in der gewählten Produktion fehlen.[5] Auch das Konstrukt ‚fit2_Musik' (Zusammenspiel von Orchester, Solisten und Chor) bereitet kaum Schwierigkeiten, obwohl auch hier explizit nach musikalischen Aspekten gefragt wurde. Im Gegensatz zu den Konstrukten ‚musikalische Leistung des Orchesters', ‚Leistung des Dirigenten' bzw. ‚musikalische Leistung des Chores' kommt es an dieser Stelle jedoch auf die Stimmigkeit an, zu deren Beantwortung weniger Wissen als vielmehr eine emotionale Einschätzung notwendig sind. Die geringe Nutzung der

[1] Trotz mitunter leicht unterschiedlichem Gebrauch der Begriffe ‚Indikator' und ‚Item' sollen diese im Folgenden synonym verwendet werden (vgl. Schnell et al., 2005, S. 152).
[2] Items dieses Konstruktes sind bspw. ‚Unterschiede in der Lautstärke wurden vom Orchester sehr gut herausgearbeitet' oder ‚Der Dirigent hatte das Orchester spürbar ‚gut im Griff'.
[3] Beispielhafte Items sind ‚Der Chor brachte Unterschiede in der Lautstärke sehr gut zur Geltung' oder ‚Der Chor überzeugte durch klangliche Brillanz'.
[4] Eine Ausnahme bildet das Item ‚Papageno sang bei seinen Soli mit einem hohen Nuancenreichtum', das von 7.6% der Befragten nicht beantwortet wurde.
[5] Im Gegensatz dazu zeigt der Chor, dessen musikalische Leistung ebenfalls relativ häufig *nicht beurteilt* werden konnte, in der bewerteten Aufführung durchaus Bühnenpräsenz. Diese ist jedoch, sowohl von ihrer Häufigkeit als auch Intensität her, in keiner Weise mit der der Solisten zu vergleichen.

Antwortmöglichkeiten „habe ich nicht beachtet“ bzw. „kann ich nicht beantworten“ überraschen bei diesem Konstrukt also nicht.

Ebenfalls auffallend bei genauer Betrachtung der nicht beantworteten Items ist die Dominanz der Antwortmöglichkeit „kann ich nicht beurteilen“, die – wenn auch schwächer – schon bei Renz (2006) und Boerner et al. (in Druck) – zu beobachten war. Demnach wurden nahezu sämtliche Dimensionen, die die Items zu operationalisieren versuchen, wahrgenommen und zur Beurteilung der künstlerischen Qualität einer Opernaufführung herangezogen, doch das hierzu erforderliche Fachwissen bzw. die Erfahrung fehlten nach Selbsteinschätzung der Teilnehmer in einigen Fällen.

Eine Ausnahme stellt das Konstrukt ‚szenische Leistung des Chores’ dar, dessen Indikatoren sehr viel häufiger *nicht beachtet* wurden als *nicht beurteilt* werden konnten. Ob dies am szenisch wenig beitragenden und bewegungsarmen Auftreten des Chores liegt, das dieser in der bewerteten Aufführung hatte, kann im Rahmen der vorliegenden Arbeit nicht abschließend geklärt werden. Eine weitere mögliche Erklärung für die verhältnismäßig häufige Nichtbeachtung des Chores in szenischer Hinsicht könnte sein, dass dieser in der gewählten Inszenierung bei nahezu sämtlichen Auftritten von einer Gruppe behinderter Menschen begleitet wurde, die das Gefolge der Alten und Kranken des Weisen Sarastro darstellten und vermutlich einen Teil der Aufmerksamkeit des Publikums auf sich zog.[1] Ein weiterer Grund könnte sein, dass der Chor in szenischer Hinsicht von Teilen des Publikums der Ausstattung bzw. dem Bühnenbild zugerechnet wird, da er in der gewählten Inszenierung wesentlich stärker im Hintergrund agierte als die Solisten. Dies kann mit Hilfe der Interkorrelation der Skalen sowie anhand von Faktorenanalysen geklärt werden (vgl. Abschnitte 4.2.2 und 4.2.3).

Die eher geringe Bedeutung des Chores für die künstlerische Qualität einer Opernaufführung wird auch durch die Frage nach der Wichtigkeit verschiedener Akteure für eine Opernaufführung allgemein, d.h. unabhängig von der bewerteten Aufführung, bestätigt, die am Schluss des Fragebogens gestellt wurde (siehe Tabelle 6).

[1] Die Nichtbeachtung des Chores wäre somit abhängig von der jeweiligen Produktion, was in weiteren Studien anhand unterschiedlicher Produktionen detailliert zu untersuchen wäre.

Tabelle 6: Wichtigkeit der verschiedenen Akteure für die künstlerische Qualität einer Opernaufführung

in % der gültigen Antworten	sehr wichtig	ziemlich wichtig	teils, teils	eher unwichtig	sehr unwichtig	Mittel-wert	Median
Orchester	88.2	11.8	--	--	--	4.88	5.00
Solisten	94.2	5.0	--	--	--	4.95	5.00
Chor	59.7	27.7	10.1	2.5	--	4.45	5.00
Dirigent	85.7	13.4	0.8	--	--	4.85	5.00
Regisseur	76.5	19.3	4.2	--	--	4.72	5.00
Bühnenbildner	52.9	42.9	3.4	0.8	--	4.48	5.00
Techniker	48.7	44.5	5.9	0.8	--	4.41	4.00
Beleuchter	50.4	42.0	5.9	1.7	--	4.41	5.00

Zusammenfassend lässt sich bisher sagen, dass die Beurteilung der musikalischen Leistung des Ensembles deutlich mehr Teilnehmern der Befragung Schwierigkeiten bereitet als die Beurteilung der szenischen, d.h. sichtbaren, Leistung. Dies kann mit dem musikalischen Wissen und der Erfahrung begründet werden, die zur Beurteilung dieser Aspekte benötigt werden und bei einigen Befragten nach eigener Einschätzung nicht vorhanden waren. Indikatoren der szenischen Leistung hingegen sprechen eher die emotionale Einschätzung des Publikums, die Beurteilung der Stimmigkeit sowie das ästhetisches Empfinden an und können damit von der überwiegenden Mehrheit des Publikums beantwortet werden (Renz, 2006, S. 30ff.; Boerner et al., in Druck).[1] Eine Ausnahme stellt das Konstrukt ‚musikalische Leistung des Solisten' (MSo) dar, dessen Items trotz fehlenden Wissens beantwortet werden. Dies ist auf die ständige szenische Präsenz des Solisten auf der Bühne zurückzuführen, die eine Bewertung seiner musikalischen Leistung offensichtlich erleichtert. Für weitere Erkenntnisse wären Studien mit der Befragung des Publikums mehrerer unterschiedlicher Produktionen notwendig, was im Rahmen dieser Arbeit aus Zeitgründen nicht möglich ist. Sollte sich bei der Datenanalyse herausstellen, dass die Konstrukte der Solisten sowie des Chores, die das analytische Modell Boerners (2002) in eine musikalische und eine szenische Funktion trennt, vom Publikum nicht eindeutig separat wahrgenommen werden, könnte dies eine mögliche

[1] Items dieser Konstrukte sind u.a.: ‚Ich konnte dem Chor durch sein Auftreten und Aussehen seine Rolle voll abnehmen', ‚Papagenos Kostüm passte sehr gut zu seiner Rolle' und ‚Das äußere Erscheinungsbild der Chorsänger (Kostüm und Maske) war ästhetisch sehr befriedigend'.

Erklärung darstellen.[1] Näheres wird die Untersuchung der Interkorrelation der Skalen zeigen, die in Abschnitt 4.2.2 vorgenommen werden wird.

Bei Betrachtung der Unterschiede in der Nutzung der Antwortmöglichkeiten „habe ich nicht beachtet“ bzw. „kann ich nicht beurteilen“ getrennt nach Laien und Experten[2] fällt auf, dass knapp 38% aller Items, die die Antwortmöglichkeiten „habe ich nicht beachtet“ bzw. „kann ich nicht beurteilen“ boten, von mindestens 5% der Laien nicht beantwortet wurden, jedoch nur knapp 7% dieser Items von mindestens 5% der Experten. Die Vermutung, der Grund für die Nichtbeantwortung einiger Items liege an fehlendem Wissen bzw. mangelnder Erfahrung, kann somit bestätigt werden.

Tabelle 7 zeigt die prozentualen Häufigkeiten nicht beantworteter Items anhand ausgewählter Konstrukte getrennt nach Laien und Experten.[3] Zur besseren Übersicht sind die Items fett markiert, die von mindestens 5% der jeweils Befragten nicht beantwortet werden konnten.

Tabelle 7: Prozentuale Häufigkeit der nicht beantworteten Items – ausgewählte Konstrukte (Laien – Experten)

Kennziffer des Items	fehlend (in Prozent aller gültigen Antworten)	Laien	Experten
Musikalische Leistung des Chores			
MC2	**5.0**	**10.5**	0.0
mMC10	**5.0**	**10.6**	0.0
MC8	**5.0**	**10.6**	0.0
MC7	**5.0**	**10.6**	0.0
Szenische Leistung des Chores			
vuSC1	1.6	3.6	0.0
vSC2	1.7	3.6	0.0
vSC3	2.5	**5.4**	0.0
vuSC4	3.3	**7.2**	0.0

[1] Betroffen wären also die Konstrukte ‚musikalische Leistung des Solisten’, ‚szenische Leistung des Solisten’ sowie ‚Zusammenspiel der Solisten’ bzw. ‚musikalische Leistung des Chores’, ‚szenische Leistung des Chores’ sowie ‚Zusammenspiel innerhalb des Chores’.
[2] Die Bildung des Indexes ‚Expertentum’ wird im nächsten Abschnitt erläutert.
[3] Eine Aufstellung der Häufigkeiten nicht beantworteter Items für sämtliche Items und Konstrukte getrennt nach Laien und Experten findet sich in Anhang K.

(Fortsetzung) Kennziffer des Items	fehlend (in Prozent aller gültigen Antworten)	Laien	Experten
Szenische Leistung des Chores			
SC6	**5.0**	**10.7**	0.0
Stimmigkeit zwischen Musik und Szene			
vfit1_1	**6.0**	**8.9**	3.6
rfit1_2	3.5	3.6	1.8
fit1_3	3.4	3.6	3.6
fit1_6	2.6	3.6	1.8
fit1_7	2.6	1.8	3.8
vfit1_8	2.6	1.8	3.7

Anmerkungen: Die Formulierungen der Items können Anhang A entnommen werden.

Schwierigkeiten, die Items zu beantworten, hatten auch die Experten bei den Konstrukten ‚musikalische Leistung des Orchesters' (MO), ‚Zusammenspiel innerhalb des Orchesters' (fit3_O) und ‚musikalische Leistung des Dirigenten' (MD), die für den Fragebogen der vorliegenden Untersuchung zusammengefasst wurden (vgl. Abschnitt 3.1.2), sowie beim Konstrukt ‚Stimmigkeit zwischen Musik und Szene' (fit1). Um einen Fragebogen zu erhalten, der die Beurteilung der künstlerischen Qualität einer Opernaufführung durch das Publikum vollständig abbildet, muss bei dessen Weiterentwicklung demnach eine weitere Vereinfachung der genannten Konstrukte in Betracht gezogen werden.

Wie zu erwarten war, wählten die Experten eher die Antwortmöglichkeit „kann ich nicht beurteilen" als „habe ich nicht beachtet", wenn sie ein Item nicht inhaltlich beantworten konnten. Eine ähnliche Tendenz ist – wenn auch weniger deutlich – bei den Laien sichtbar (siehe Anhang J). Die Experten nehmen demnach nahezu alle Indikatoren, nach denen im Fragebogen gefragt wurde, in ihre Beurteilung der künstlerischen Qualität auf und urteilen somit differenzierter als Laien. So wurden bspw. alle Items zur szenischen Leistung des Chores von allen Experten beantwortet, wohingegen ein Teil der Laien die Items hauptsächlich deshalb nicht beantwortete, weil er die Indikatoren *nicht beachtet* hatte. Jedoch mangelt es auch bei Teilen der Experten nach eigener Einschätzung an Fachwissen und Erfahrung, so dass auch unter diesen nicht alle Items vollständig beantwortet werden konnten. Eine weitere Erklärungsmöglichkeit für die fehlenden Werte auch unter den Experten ist, dass

die Beantwortung einiger Items möglicherweise auch mit entsprechendem Wissen und vorhandener Erfahrung Schwierigkeiten bereitet.

Welche Schlussfolgerungen können nun aus der Analyse der prozentualen Häufigkeiten der Antwortkategorien „habe ich nicht beachtet“ bzw. „kann ich nicht beurteilen“ gezogen werden? Zusammenfassend lässt sich konstatieren, dass die Bewertung der musikalischen Aspekte einer Opernaufführung sowohl den Laien als auch den Experten deutlich mehr Schwierigkeiten bereitet als die Bewertung der szenischen Aspekte. Es wird vermutet, dass dies an fehlendem Wissen bzw. mangelnder Erfahrung liegt, die zur Beantwortung der Items notwendig sind (Renz, 2006). Dies wird dadurch bestätigt, dass die Anzahl der nicht beantworteten Items bei den Experten deutlich unter der der Laien liegt. Als erstes Ergebnis lässt sich schließen, dass zur Beurteilung der künstlerischen Qualität einer Opernaufführung eine Vielzahl von Konstrukten herangezogen wird, die durch das analytische Modell Boerners (2002) bzw. den weiterentwickelten Fragebogen sehr gut abgebildet werden. Dabei differenzieren die Experten in größerem Ausmaß als die Laien.

Dieses Ergebnis wird durch die genaue Betrachtung der Verteilung der beiden Antwortkategorien bestätigt. Sowohl die Laien als auch die Experten wählen häufiger die Möglichkeit „kann ich nicht beurteilen“, wenn sie sich nicht in der Lage sehen, ein Item inhaltlich zu beantworten, d.h. es wird im Fragebogen mitunter nach Kriterien gefragt, die zwar im Analyseraster der Zuschauer, jedoch außerhalb der selbsteingeschätzten Urteilskompetenz liegen. Die Dominanz von „kann ich nicht beurteilen“ über „habe ich nicht beachtet“ ist bei den Experten noch ausgeprägter, was die Vermutung hinsichtlich der größeren Differenzierung durch die Experten weiter bestätigt. Gleichzeitig fällt auf, dass auch von den Experten nicht sämtliche Items beantwortet werden können; neben Wissen und Erfahrung scheinen demnach weitere Kenntnisse zur Beantwortung der Items notwendig zu sein.

Durch die erfolgte deskriptive Auswertung ergab sich neben Erkenntnissen zum Ausmaß der Übereinstimmung zwischen der Struktur der empirisch wahrgenommenen Aufführungsqualität und dem hierarchischen Modell (Boerner, 2002) ein erster Eindruck davon, welche musikalischen und szenischen Aspekte einer Opernaufführung vom Publikum beachtet und unterschieden werden (Untersuchungsfrage 1a). Welche Aspekte jedoch tatsächlich unterschieden werden, wird im Rahmen von Faktorenanalysen in Abschnitt 4.2.3 untersucht.

4.1.4 Bildung des Indexes ‚Expertentum'

Anders als bei Boerner et al. (in Druck), die den Kenntnisstand des Publikums anhand einer einzigen Variablen erhoben hatten[1], wurde die latente Variable ‚Expertentum' in der vorliegenden Untersuchung mit Hilfe mehrerer Items erhoben.

Diese Vorgehensweise ist dann zu wählen, „wenn entweder ein einzelner Indikator die interessierende Dimension nicht mit ausreichender Genauigkeit misst oder die Begriffe einer sozialwissenschaftlichen Theorie mehrere Dimensionen ansprechen" (Schnell et al., 2005, S. 166). In diesem Fall werden mehrere Indikatoren zu einer neuen Variablen zusammengefasst, die dann als ‚Index' bezeichnet wird.

Um eine breitere Basis für die Selbsteinschätzung zu erhalten und damit ‚Expertentum' möglicherweise genauer zu erfassen, wurden mehrere Indikatoren zum Index ‚Expertentum' zusammengefasst. Neben dem Item ‚Wie groß ist Ihre Erfahrung in Bezug auf die Oper?' sind dies die folgenden: ‚Wie oft pro Jahr besuchen Sie eine Opernaufführung (im Durchschnitt)?', ‚Beherrschen Sie ein Instrument?' und ‚Wie gut haben Sie das aufgeführte Stück bereits gekannt?'. Die beiden erstgenannten Items decken dabei die Dimension ‚Erfahrung' ab, wohingegen sich die restlichen zwei auf die Dimension ‚Fachkenntnis' beziehen, die jeweils übliche Dimensionen zur Erfassung von Expertentum darstellen (Chi et al., 1988; Ericsson & Smith, 1991; Ericsson et al., 1993). Hierbei ist zu erwähnen, dass die Auswahl der Items und damit der Dimensionen, die den Index bilden, eine rein theoretische Arbeit darstellt und es keine objektiven Kriterien gibt, ob alle relevanten Dimensionen in den Index eingehen bzw. ob alle eingehenden Dimensionen von Relevanz sind (Schnell et al., 2005, S. 169).

Nach der Bestimmung der vier Items, die den Index ‚Expertentum' bilden, stellt sich die Frage, auf welche Weise die Werte der verschiedenen Items zu einem Index sinnvoll zusammengesetzt werden. Aus Gründen der Übersichtlichkeit soll der Index ‚Expertentum' für diese Untersuchung nur zwei Ausprägungen einnehmen können: Laie und Experte. Zu diesem Zweck wurden die vier gewählten Items je nach Ausprägung mit verschiedenen Punktwerten versehen. Da in dieser Arbeit angenommen wird, dass das Item ‚Erfahrung in Bezug auf die Oper' Expertentum am unmittelbarsten misst und die weiteren Items die

[1] Die Formulierung des Items lautete: ‚Wie würden Sie Ihre Erfahrung in Bezug auf Musiktheater einschätzen?'. Die Antwortmöglichkeiten waren: „Ich bin absoluter Laie", „Ich bin Laie, habe aber etwas Musiktheatererfahrung", „Musiktheater gehört zu meinen ‚Hobbies', ich gehe häufiger in Aufführungen", „Ich würde mich schon als ‚Halbprofi' bezeichnen" sowie „Experte,-in".

Selbsteinschätzung der Erfahrung objektiv bestätigen sollen, konnten durch dieses Item maximal 10 Punkte und durch die anderen drei jeweils maximal 5 Punkte erreicht werden. Die Punkte aller vier Dimensionen wurden anschließend addiert, so dass maximal 25 Punkte erreicht werden konnten. Folgende Häufigkeitsverteilung ergab sich aus der beschriebenen Vorgehensweise (siehe Tabelle 8):

Tabelle 8: Häufigkeitsverteilung der Punkte zur Bildung des Indexes ‚Expertentum'

		Häufigkeit	Prozent	gültige Prozente	kumulierte Prozente
gültig	6	1	.8	.9	.9
	7	2	1.7	1.8	2.6
	8	4	3.3	3.5	6.1
	9	5	4.2	4.4	10.5
	10	8	6.7	7.0	17.5
	11	13	10.8	11.4	28.9
	12	13	10.8	11.4	40.4
	13	11	9.2	9.6	50.0
	14	15	12.5	13.2	63.2
	15	7	5.8	6.1	69.3
	16	8	6.7	7.0	76.3
	17	11	9.2	9.6	86.0
	18	8	6.7	7.0	93.0
	19	5	4.2	4.4	97.4
	20	1	.8	.9	98.2
	23	1	.8	.9	99.1
	24	1	.8	.9	100.0
	gesamt	114	95.0	100.0	
fehlend		6	5.0		
gesamt		120	100.0		

Anschließend wurde der Index in ‚Laie' (0 – 13 Punkte) und ‚Experte' (14 – 25 Punkte) eingeteilt, so dass sich unter den Befragten genauso viele Laien wie Experten befinden (jeweils n = 57) und Vergleiche beider Gruppen die Annahme gleicher Gruppengrößen nicht verletzen. Zu bemerken ist abschließend, dass auf die beschriebene Weise auch opernerfahrene Nicht-Professionelle mit Kenntnissen der aufgeführten Oper als ‚Experten' bezeichnet werden, so dass Vergleiche mit der Literatur, in der es sich bei ‚Experten' zumeist um Künstler oder professionelle Kritiker handelt (z.B. Koebner, 1993; Boorsma & van Maanen, 2003; Eversmann, 2004) mit Vorsicht zu erfolgen haben.

4.2 Reliabilität und Validität des Fragebogens

Da sich die Konstrukte, die in dieser Untersuchung interessieren, nicht direkt beobachten lassen und zusätzlich jede Messung potenziell mit Ungenauigkeiten verbunden ist, ist bei der

Messung der Kriterien, die ein Opernpublikum zur Bewertung künstlerischer Qualität heranzieht, mit zwei Arten von Fehlern zu rechnen: Erstens können systematische Fehler auftreten, die dann entstehen, wenn eine Skala die falsche Größe misst. In diesem Fall wäre die Skala nicht valide, d.h. gültig. Die zweite Art von Fehlern, die auftreten kann, sind Zufallsfehler, die bei der Erhebung von Daten im Rahmen von Befragungen nahezu unvermeidlich sind. Brosius (2006) nennt Ungenauigkeiten beim Beobachten der Werte sowie Interpretationsspielräume im Rahmen von Befragungsverfahren als häufigste Quellen von Zufallsfehlern (Brosius, 2006, S. 799).

4.2.1 Reliabilitätsanalysen

Als Reliabilität wird das Ausmaß definiert, „in dem wiederholte Messungen eines Objektes mit einem Messinstrument die gleichen Werte liefern“ (Schnell et al., 2005, S. 151).[1] Da nach der Messung jedoch lediglich die beobachteten Werte bekannt und die wahren Werte bzw. die Größe der Fehler nach wie vor unbekannt sind, ist die exakte Reliabilität einer Skala nur schwer zu bestimmen. Eine in der Literatur sehr häufig verwendete Methode ist die Analyse der internen Konsistenz[2] durch die Berechnung von Cronbachs Alpha-Koeffizienten. Dazu wird das Messinstrument auf so viele Weisen wie möglich in zwei Hälften gespalten und der Mittelwert aller möglichen Korrelationen der beiden Testhälften gebildet. Cronbachs Alpha kann Werte zwischen 0 und 1 annehmen, wobei in der Literatur häufig Werte von mindestens 0.8 gefordert und Werte ab 0.7 akzeptiert werden (Brosius, 2006, S. 800).

Zur Messung der Güte des im Rahmen dieser Arbeit weiterentwickelten Fragebogens wird aus zuvor erläuterten Gründen zunächst die Reliabilität der vom analytischen Modell Boerners (2002) theoretisch angenommenen sowie der zur Erfassung des Gesamterlebnisses des Abends in der Oper neu gebildeten Konstrukte überprüft, d.h. es wird getestet, ob diese reliable Skalen bilden. Tabelle 9 gibt Auskunft über Reliabilität, d.h. Cronbachs Alpha der

[1] Zur Berechnung der Reliabilität wird der Quotient der Varianz der wahren Werte und der Varianz der beobachteten Werte gebildet, so dass die Reliabilität im Umkehrschluss das Quadrat der Korrelation zwischen den beobachteten und den wahren Werten ist (Schnell et al., 2005, S. 151).

[2] Eine Möglichkeit der Schätzung der Reliabilität ist, ein Messinstrument aus mehreren Indikatoren zu bilden, die möglichst alle dieselbe Dimension bzw. Größe messen. Da angenommen werden kann, dass sich die zufälligen Fehler bei jedem Item in unterschiedlicher Weise über die verschiedenen Befragten verteilen, erlaubt es der Zusammenhang zwischen wahrer Varianz, die bei allen Items gleich zum Ausdruck kommt, und Fehlervarianz, Rückschlüsse über ihre jeweiligen Anteile zu ziehen und so die Reliabilität der Skala zu bestimmen. So genannte ‚Maße der internen Konsistenz’ versuchen auf diese Weise Aussagen darüber zu machen, in welchem Umfang die Indikatoren dasselbe Konstrukt messen und dienen als Basis für Reliabilitätsschätzungen.

Skalen, die Anzahl verwendeter Items[1] sowie die Kennziffern der Items, aus denen die jeweiligen Skalen gebildet werden.

Tabelle 9: Reliabilität der gebildeten Skalen

Skala	Reliabilität	Anzahl verwendeter Items (Anzahl ursprünglicher Items)	Kennziffern der verwendeten Items
Gesamterlebnis Opernabend	.94	4 (5)	mGE1, mGE2, mrGE3, mGE4
künstlerische Qualität	.90	3 (4)	mGQ1, mGQ2, vNG2
MOrchester / fit3_ Orchester / Dirigent	.97	10 (11)	mMO13, vMO2, vuMO3, vMO5, MO10, mfit3O11, fit3O5, MD1, vMD7, MD13
MSolist	.91	5 (6)	MS1, MS5, vMS6, vuMS7, MS10
MChor	.96	4 (4)	MC2, mMC10, MC8, MC7
SSolist	.86	4 (5)	SS1, vSS4, SS5, vSS8
SChor	.94	3 (5)	vuSC1, vSC2, vSC3,
SAusstattung	.91	4 (4)	vSA13, mSA16, vSA15, mSA17
fit3_Chor	.81	3 (3)	fit3C1, vufit3C2, vfit3C3
fit3_Solisten	.79	3 (4)	vfit3S1, rfit3S3, ufit3S5
fit2_Musik	.82	5 (5)	fit2_2, vfit2_3, rfit2_6, vfit2_8, vrfit2_9
fit3_Ausstattung und fit2_Szene	.95	8 (9)	fit3A1, vfit3A2, fit3A4, fit3A5, vfit2S_1, fit2S_2, vrfit2S_4, vfit2S_3
fit_1	.92	5 (6)	vfit1_1, fit1_3, fit1_6, fit1_7, vfit1_8
Komplexität	.72	3 (3)	rKA1_k, rKA2_k, rKA3_k
Neuartigkeit	.86	3 (3)	rKA9_v, ruGA6_v, vrKA7_v
Identifikation	.91	3 (3)	mII1_i, mII2_i, mII3_i
Empathie	.92	3 (3)	mEE1, mEE2, mEE3
persönliche Assoziationen	.71	3 (4)	mEA1, mEA3, mEA4
Genussfaktor	.90	3 (3)	mEP1, mEP2, mEP3_2

[1] Es wurden jeweils die Items mit der geringsten Trennschärfe, d.h. der Korrelation zwischen dem jeweiligen Item und der Gesamtskala, die sich unter Ausschluss des entsprechenden Items ergäbe, entfernt (Brosius, 2006, S. 802). Auch bei geringer Trennschärfe muss eine Skala jedoch aus mindestens drei Items bestehen.

(Fortsetzung) Skala	Reliabilität	Anzahl verwendeter Items (Anzahl ursprünglicher Items)	Kennziffern der verwendeten Items
Ich-Beteiligung	.90	3 (3)	mEI1, vKA4_EI, mEI3
vorübergehende Gefühlsregung	.84	3 (3)	mES1, mES2, mES3
Gefühl des Bewegt-Seins	.88	3 (3)	rNG7, vNG6, mrEM3
situative Rahmenbedingungen	.72	8 (8)	mRR1, mRR2, mRR3, mRR4, mRR5, mRR6, mRR7, mRR8

Anmerkungen: MOrchester = musikalische Leistung des Orchesters, fit3_Orchester = Zusammenspiel innerhalb des Orchesters, MSolist = musikalische Leistung des Solisten, MChor = musikalische Leistung des Chores, SSolist = szenische Leistung des Solisten, SChor = szenische Leistung des Chores, SAustattung = Ausstattung (Einzelaspekte), fit3_Chor = Zusammenspiel innerhalb des Chores, fit3_Solisten = Zusammenspiel der Solisten untereinander, fit2_Musik = Stimmigkeit innerhalb der musikalischen Dimension, fit3_Ausstattung = Stimmigkeit der Ausstattung, fit2_Szene = Stimmigkeit innerhalb der szenischen Dimension, fit1 = Stimmigkeit zwischen der musikalischen und der szenischen Dimension. Die Formulierungen der Items können Anhang A entnommen werden.

Nachdem es bei Renz (2006) aufgrund der – bei einigen Konstrukten – sehr hohen Anzahl nicht beantworteter Items nicht möglich war, Skalen zu sämtlichen Konstrukten des analytischen Modell Boerners (2002) zu bilden[1], ist dies in der vorliegenden Studie ohne Probleme möglich, da kein Item von mehr als 10% der Befragten nicht beantwortet wurde. Die Modifikation der Konstrukte zur Erfassung künstlerischer Qualität in der Oper (vgl. Abschnitt 3.1.2) scheint demnach gelungen zu sein.

Unter Verwendung des häufig herangezogenen Kriteriums der Literatur, nach dem Werte von Cronbachs Alpha ab 0.8 akzeptabel sind, können die gebildeten Skalen zur Erfassung der künstlerischen Qualität als gut bis sehr gut bezeichnet werden.[2] Es kann davon ausgegangen werden, dass die theoretisch angenommenen Konstrukte (Boerner, 2002) durch die jeweiligen Items reliabel gemessen werden. Wie Anhang D entnommen werden kann, verbessern sich die Reliabilitäten gegenüber Moser (2005) und Renz (2006) deutlich und sind ähnlich hoch wie bei Boerner et al. (in Druck). Auch sämtliche neu gebildeten Skalen zur Erfassung der Determinanten des Gesamturteils des Publikums über einen Abend in der Oper (‚Gesamterlebnis Opernabend') erreichen Werte von Cronbachs Alpha über 0.7 und können als reliabel bezeichnet werden. Dies ist umso bemerkenswerter, als dass mehrere Skalen nur

[1] Betroffen bei der Aufführung „Aida" waren die Konstrukte ‚musikalische Leistung des Chores', ‚Leistung des Dirigenten', ‚Zusammenspiel innerhalb des Orchesters' sowie ‚Zusammenspiel innerhalb des Chores'.
[2] Einzige Ausnahme stellt die Skala ‚Zusammenspiel der Solisten untereinander' dar, deren Reliabilität knapp unter 0.8 liegt (Cronbachs Alpha = 0.79).

aus drei Items gebildet werden und Cronbachs Alpha mit der Anzahl der Items generell steigt.[1]

4.2.2 Interkorrelation der Skalen

Zur Prüfung der Validität des Fragebogens (vgl. Abschnitt 4.2) werden im Folgenden die Interkorrelationen der gebildeten Skalen[2] berechnet, womit die Beziehungen zwischen den Skalen und somit deren Unterscheidbarkeit getestet werden. Ein geeignetes Maß sowohl für die Stärke als auch die Richtung des Zusammenhangs zwischen zwei Skalen stellt Pearsons Korrelationskoeffizient r dar, welcher die Stärke eines (linearen) Zusammenhangs in einer Zahl zwischen -1 und +1 ausdrückt (siehe Tabelle 10).[3] Zu beachten ist jedoch, dass selbst eine sehr starke oder sogar perfekte Korrelation keinen Beleg für einen kausalen Zusammenhang darstellt (Brosius, 2006, S. 519).[4]

[1] Die Reliabilität einer Skala steigt deshalb mit jedem addierten Item, weil sich auf diese Weise die zufälligen Messfehler gegenseitig eliminieren, die dann im Mittel den Wert 0 haben (Brosius, 2006, S. 800).
[2] Eine Übersicht, aus welchen Items die verschiedenen Skalen gebildet wurden, bietet Tabelle 9 in Abschnitt 4.2.1.
[3] Ein positiver Wert zeigt dabei einen positiven Zusammenhang an, ein negativer Wert deutet auf einen negativen Zusammenhang hin; der Betrag ist umso größer, je stärker der Zusammenhang ist.
[4] Für eine ausführliche Kausalitätsdiskussion jüngerer Zeit vgl. Salmon (1998) und Pearl (2000).

Tabelle 10: Interkorrelation der Skalen

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	1																						
2	.88**	1																					
3	.68**	.72**	1																				
4	.66**	.67**	.67**	1																			
5	.58**	.59**	.70**	.59**	1																		
6	.72**	.73**	.71**	.78**	.65**	1																	
7	.44**	.48**	.29*	.38**	.35**	.63**	1																
8	.63**	.62**	.31*	.41**	.32**	.57**	.62**	1															
9	.49**	.59**	.67**	.59**	.64**	.61**	.30*	.35**	1														
10	.63**	.67**	.71**	.72**	.49**	.57**	.20	.40**	.72**	1													
11	.55**	.58**	.64**	.53**	.43**	.48**	.21	.29*	.64**	.68**	1												
12	.53**	.52**	.32**	.39**	.37**	.63**	.69**	.84**	.40**	.30*	.32**	1											
13	.73**	.73**	.51**	.56**	.44**	.73**	.68**	.73**	.55**	.50**	.47**	.80**	1										
14	-.53**	-.49**	-.47**	-.47**	-.50**	-.42**	-.13	-.34**	-.49**	-.49**	-.40**	-.33**	-.39**	1									
15	-.69**	-.65**	-.26*	-.44**	-.29*	-.56**	-.57**	-.78**	-.25*	-.29*	-.27*	-.75**	-.69**	.47**	1								
16	.42**	.42**	.27*	.47**	.36**	.64**	.51**	.51**	.33**	.30*	.09	.49**	.53**	-.32**	-.47**	1							
17	.71**	.75**	.66**	.64**	.55**	.71**	.38**	.50**	.58**	.59**	.48**	.39**	.64**	-.51**	-.45**	.61**	1						
18	.42**	.53**	.37**	.39**	.37**	.54**	.44**	.54**	.48**	.36**	.34**	.45**	.55**	-.28*	-.35**	.59**	.72**	1					
19	.91**	.83**	.63**	.67**	.60**	.70**	.44**	.65**	.45**	.56**	.48**	.53**	.71**	-.49**	-.67**	.48**	.70**	.46**	1				
20	.88**	.87**	.64**	.69**	.55**	.76**	.51**	.67**	.52**	.58**	.50**	.55**	.73**	-.43**	-.68**	.49**	.70**	.56**	.87**	1			
21	.53**	.64**	.41**	.53**	.24*	.56**	.51**	.61**	.39**	.39**	.34**	.44**	.61**	-.20	-.56**	.42**	.53**	.46**	.57**	.75**	1		
22	.73**	.79**	.60**	.63**	.49**	.69**	.49**	.65**	.54**	.59**	.51**	.58**	.73**	-.38**	-.61**	.40**	.62**	.51**	.69**	.80**	.72**	1	
23	.21	.21	.38**	.14	.17	.19	.10	.11	.46**	.34**	.36**	.16	.15	-.36**	-.09	.17	.25*	.24*	.10	.14	-.03	.17	1

Anmerkungen: ** $p \leq 0.01$, * $p \leq 0.05$; n = 71 (listenweiser Fallausschluss). 1 = Gesamterlebnis Opernabend, 2 = künstlerische Qualität, 3 = musikalische Leistung und Zusammenspiel des Orchesters sowie Leistung des Dirigenten, 4 = musikalische Leistung des Solisten, 5 = musikalische Leistung des Chores, 6 = szenische Leistung des Solisten, 7 = szenische Leistung des Chores, 8 = Ausstattung (Einzelaspekte), 9 = Zusammenspiel innerhalb des Chores, 10 = Zusammenspiel der Solisten untereinander, 11 = Stimmigkeit innerhalb der musikalischen Dimension, 12 = Stimmigkeit innerhalb der szenischen Dimension, 13 = Stimmigkeit zwischen der musikalischen und der szenischen Dimension, 14 = Komplexität, 15 = Neuartigkeit, 16 = Identifikation, 17 = Empathie, 18 = persönliche Assoziationen, 19 = Genussfaktor, 20 = Ich-Beteiligung, 21 = vorübergehende Gefühlsregung, 22 = Gefühl des Bewegt-Seins, 23 = situative Rahmenbedingungen.

Die Korrelationskoeffizienten liegen zwischen r = -.78 und r = .91 und sind zum Großteil hoch signifikant ($p \leq 0.01$). Zur inhaltlichen Interpretation der Skalen zur künstlerischen Qualität erscheint die getrennte Betrachtung der hierarchischen Ebenen des analytischen Modell Boerners (2002) sinnvoll.

Auf der dritten Ebene fallen die starken Korrelationen des Konstruktes ‚musikalische Leistung sowie Zusammenspiel des Orchesters und Leistung des Dirigenten'[1] mit der ‚musikalischen Leistung des Solisten' (r = .67, $p \leq 0.01$), der ‚musikalischen Leistung des Chores' (r = .70, $p \leq 0.01$), dem ‚Zusammenspiel innerhalb des Chores' (r = .67, $p \leq 0.01$) sowie dem ‚Zusammenspiel der Solisten untereinander' (r = .71, $p \leq 0.01$) auf. Offensichtlich ist die Wahrnehmung der musikalischen Aspekte weniger differenziert als zunächst angenommen, zumal ebenfalls die Konstrukte ‚musikalische Leistung des Solisten', ‚musikalische Leistung des Chores', ‚Zusammenspiel innerhalb des Chores' und ‚Zusammenspiel der Solisten untereinander' jeweils stark korrelieren. Inwieweit die ‚physisch' unterscheidbaren Konstrukte Orchester, Solisten und Chor in musikalischer Hinsicht in der Wahrnehmung des Publikums unterschieden werden, wie vom analytischen Modell (Boerner, 2002) vermutet, muss eine anschließende Faktorenanalyse zeigen.

Da der Solist sowie der Chor durch das analytische Modell (Boerner, 2002) in zwei Konstrukte getrennt werden (musikalische Leistung/ szenische Leistung), obwohl diese physisch jeweils ein Ganzes bilden, muss geprüft werden, wie die Skalen jeweils zusammenhängen. Es ergeben sich widersprüchliche Erkenntnisse: So korrelieren die Skalen ‚musikalische Leistung des Solisten' und ‚szenische Leistung des Solisten' sehr stark (r = .78, $p \leq 0.01$), wie auch die ‚musikalische Leistung des Solisten' mit dem ‚Zusammenspiel der Solisten untereinander' (r = .72, $p \leq 0.01$) und das ‚Zusammenspiel der Solisten untereinander' mit der ‚szenischen Leistung des Solisten' (r = .57, $p \leq 0.01$) korrelieren. Die Skalen überlappen sich also zu einem Teil, messen jedoch keineswegs Identisches. Nur geringe und zum Teil nicht signifikante Korrelationen bestehen hingegen zwischen den verschiedenen Skalen des Chores (vgl. Moser, 2005, S. 51).

[1] Die Zusammenlegung der vormals getrennten Konstrukte ‚musikalische Leistung des Orchesters', ‚Zusammenspiel innerhalb des Orchesters' und ‚Leistung des Dirigenten' (vgl. Abschnitt 3.1.2) wird, nach bereits erfolgter Bestätigung durch die Reliabilitätsanalysen, nun auch durch die Interkorrelation der Skalen bestätigt. Die einzelnen Konstrukte korrelieren bei getrennter Betrachtung hoch signifikant und sehr stark, so dass von einer separaten Wahrnehmung des Publikums nicht ausgegangen werden kann. Selbiges gilt für die Skalen ‚Stimmigkeit innerhalb der szenischen Dimension' und ‚Stimmigkeit der szenischen Aspekte', die im Rahmen der Validierung des Fragebogens ebenfalls zusammengefasst wurden. Eine Übersicht über Pearsons Korrelationskoeffizienten der separaten Konstrukte findet sich in Anhang L.

Auffallend stark ist die Korrelation des Konstruktes ‚szenische Leistung des Solisten' mit einer Vielzahl anderer – sowohl musikalischer als auch szenischer – Konstrukte. Die szenische Leistung des Solisten erscheint damit als eine Art ‚Generalfaktor', mit dessen Hilfe sich eine generelle Antworttendenz beschreiben lässt. Anders als bei Moser (2005) und Renz (2006) gilt dies jedoch nicht für die Szene allgemein, die sich in deren Untersuchungen jeweils als Generalfaktor zeigte (Moser, 2005, S. 52; Renz, 2006, S. 44). Die szenischen Konstrukte werden, abgesehen von der ‚szenischen Leistung des Solisten', also durchaus in der Wahrnehmung des Publikums voneinander unterschieden; die von Renz (2006) vorgenommene Ausdifferenzierung der Ausstattungseffekte (Renz, 2006, S. 18) erscheint gerechtfertigt.

Ebenfalls starke Korrelationen zeigen sich zwischen der zweiten und dritten Ebene, die im Gegensatz zu denen innerhalb der dritten Ebene der Logik des analytischen Modell Boerners (2002) entsprechen.[1] So korreliert die ‚Stimmigkeit innerhalb der musikalischen Dimension' signifikant mit der ‚musikalischen Leistung des Orchesters' ($r = .64$, $p \leq 0.01$), der ‚musikalischen Leistung des Solisten' ($r = .53$, $p \leq 0.01$), der ‚musikalischen Leistung des Chores' ($r = .43$, $p \leq 0.01$), dem ‚Zusammenspiel innerhalb des Chores' ($r = .64$, $p \leq 0.01$) sowie dem ‚Zusammenspiel der Solisten untereinander' ($r = .68$, $p \leq 0.01$) (vgl. Moser, 2005, S. 52). Nur sehr schwach bis schwach korreliert die ‚Stimmigkeit innerhalb der musikalischen Dimension' hingegen mit den szenischen Aspekten.[2] Genau umgekehrt präsentieren sich die Korrelationen der ‚Stimmigkeit innerhalb der szenischen Dimension', die – gemäß der Logik des analytischen Modells (Boerner, 2002) – stark mit den szenischen Aspekten der dritten Ebene (‚szenische Leistung des Solisten', ‚szenische Leistung des Chores' sowie ‚Szene_Ausstattung') und nur schwach mit den musikalischen Aspekten korrelieren.

Ebenfalls erwähnenswert sind die Korrelationen der ersten mit der zweiten Ebene: So korreliert die ‚Stimmigkeit zwischen der musikalischen und der szenischen Dimension' sehr stark mit der ‚Stimmigkeit innerhalb der szenischen Dimension' ($r = .80$, $p \leq 0.01$), nur mittelmäßig hingegen mit der ‚Stimmigkeit innerhalb der musikalischen Dimension' ($r = .47$, $p \leq 0.01$). Die szenische Dimension scheint für die Stimmigkeit zwischen Szene und Musik erheblich bedeutsamer zu sein; auf einen kausalen Zusammenhang kann von einer starken

[1] Sämtliche Unterkonstrukte der Dimension ‚Musik' (dritte Ebene) sind im analytischen Modell in der Skala ‚Stimmigkeit innerhalb der musikalischen Dimension' (zweite Ebene) enthalten.
[2] Eine Ausnahme, die inhaltlich nicht plausibel erscheint, bildet das Konstrukt ‚szenische Leistung des Solisten' ($r = .48$, $p \leq 0.01$), welches, wie oben gezeigt, als Generalfaktor interpretiert werden kann.

Korrelation – wie bereits erwähnt – jedoch keineswegs geschlossen werden. Darüber hinaus korreliert die ‚Stimmigkeit zwischen der musikalischen und der szenischen Dimension' hoch signifikant mittel bis stark mit sämtlichen Skalen beider Dimensionen. Dies entspricht der Logik des analytischen Modells (Boerner, 2002), in dem die ‚Stimmigkeit zwischen der musikalischen und der szenischen Dimension' die oberste Aggregatebene bildet und somit alle Teilkonstrukte beider Dimensionen subsumiert (vgl. Abschnitt 2.2.1).

Mit der Skala ‚künstlerische Qualität der Opernaufführung' korrelieren ebenfalls sämtliche Skalen des hierarchischen Modells (Boerner, 2002) signifikant positiv und zum Großteil erheblich. Die stärkste Korrelation besteht dabei mit den Skalen ‚Stimmigkeit zwischen der musikalischen und der szenischen Dimension' ($r = .73$, $p \leq 0.01$) und ‚szenische Leistung des Solisten' ($r = .73$, $p \leq 0.01$).

Bei Betrachtung der Beziehungen zwischen den Skalen zur Erfassung der Determinanten des Gesamturteils des Publikums über einen Abend in der Oper (‚Gesamterlebnis Opernabend') fallen zunächst die starken Korrelationen der Skala ‚Gesamterlebnis Opernabend' mit nahezu sämtlichen anderen Skalen in Auge. Am stärksten korreliert sie dabei mit den Skalen ‚Genussfaktor' ($r = .91$, $p \leq 0.01$), ‚Ich-Beteiligung' ($r = .88$, $p \leq 0.01$) und ‚künstlerische Qualität' ($r = .88$, $p \leq 0.01$). Ebenfalls auffällig sind die Korrelationskoeffizienten der Skalen ‚Komplexität' und ‚Neuartigkeit', die ohne Ausnahme negativ sind. Positiv und verhältnismäßig hoch hingegen sind die Korrelationen der Skalen, die die emotionale Wirkung einer Opernaufführung auf den Zuschauer erfassen – sowohl untereinander als auch mit den Skalen zur künstlerischen Qualität. Welche Teilaspekte von den Zuschauern tatsächlich unterschieden werden, wird daher im Folgenden anhand von Faktorenanalysen genauer untersucht.

4.2.3 Faktorenanalysen

Nach Betrachtung der Interkorrelation der Skalen stellen sog. Faktorenanalysen eine weitere Methode der Validierung des modifizierten Fragebogens dar. Als Faktorenanalysen bezeichnet man eine Reihe statistischer Verfahren, mit deren Hilfe sich testen lässt, ob sich unter den erhobenen Items solche Gruppen von Items befinden, denen jeweils eine unbeobachtete Hintergrundvariable (‚Faktor') zugrunde liegt (Brosius, 2006, S. 763).[1]

[1] Ähnliches, so könnte ein (durchaus berechtigter) Einwand lauten, sei doch bereits mit der Reliabilitätsanalyse versucht worden. Weil Cronbachs Alpha – ceteris paribus – umso höher ist, je stärker die einzelnen Items miteinander korrelieren, wird dieser Wert tatsächlich oftmals dazu benutzt, die Eindimensionalität von Items zu

Faktorenanalyse der ersten Ebene des analytischen Modells

Gemäß der Logik des analytischen Modell Boerners (2002, S. 92f.) werden im Folgenden die drei hierarchischen Ebenen getrennt untersucht. Tabelle 11 zeigt die Faktorladungen der ersten Ebene des Modells, bei der es um die Stimmigkeit zwischen der musikalischen und der szenischen Dimension geht (siehe Abbildung 1). Wie erwartet lässt sich nur ein Faktor extrahieren, auf den sämtliche Items mit hohen Koeffizienten (≥ .80) laden und der 77% der Varianz aufklärt. Im Vergleich zur Studie von Renz (2006) wurden damit sowohl die Faktorladungen als auch die Varianzaufklärung deutlich gesteigert (2006, S. 46), welches für die Validität des modifizierten Fragebogens spricht. Dabei wurde eine explorative Herangehensweise gewählt, bei der versucht wird, sowohl die Anzahl der Faktoren als auch die Zusammenhänge zwischen den Faktoren auf der einen und zwischen den Faktoren und den Items auf der anderen Seite aus den Daten zu errechnen.[1]

Tabelle 11: Faktorenanalyse der ersten Ebene der Oper „Die Zauberflöte“

Kennziffern der Items	Label der Items	fit1 Faktorladungen
vfit1_1	Musik – Szene	.90
fit1_3	Charakter der Musik	.89
fit1_6	unterstrichen	.80
fit1_7	Musik und Szene wie aus einem Guss	.89
vfit1_8	Unterstützung	.91
Varianzaufklärung		77%

Anmerkungen: Extraktionsmethode: Hauptkomponentenanalyse. Keine Vorgabe der Anzahl von Faktoren. Die Formulierungen der Items können Anhang A entnommen werden.

Faktorenanalyse der zweiten Ebene des analytischen Modells

Auch auf der zweiten Ebene des analytischen Modell Boerners (2002) geht es um Stimmigkeit; im Unterschied zur ersten Ebene jedoch um jene innerhalb der beiden Dimensionen. So sind „innerhalb der musikalischen Dimension […] die Potentialfaktoren

zeigen, d.h. es wird versucht zu ermitteln, ob bestimmte Items dieselbe Größe messen. Obschon es unter bestimmten Umständen möglich ist, aus den Werten von Cronbachs Alpha Hinweise auf die Dimensionalität von Items zu gewinnen, ist diese Vorgehensweise im vorliegenden Fall nicht angebracht. Auch für eine Gruppe aus mehreren Items, die mehr als eine Dimension abbilden, könnte sich ein hohes Alpha ergeben. Dies geschieht dann, wenn es mehrere Grüppchen von Items gibt, die dieselbe Dimension abbilden und daher untereinander stark korrelieren und gleichzeitig die Items verschiedener Dimensionen nicht stark korrelieren. Ebenso ist der umgekehrte Fall denkbar, d.h. ein niedriges Alpha trotz Eindimensionalität, wenn nämlich mehrere Items, die dieselbe Dimension abbilden, mit hohen Zufallsfehlern behaftet sind (Brosius, 2006, S. 801). Daher ist es zur Identifikation unbeobachteter Hintergrundvariablen notwendig, trotz erfolgter Reliabilitätsanalyse eine Faktorenanalyse durchzuführen.

[1] Eine Alternative stellen konfirmatorische Faktorenanalysen dar, mit deren Hilfe getestet wird, ob exakt spezifizierte Hypothesen über zu erwartende Ergebnisse durch die Daten gestützt werden und die daher die vorherige Festlegung der Anzahl und der Beziehungen der Faktoren sowie der Beziehungen zwischen den Faktoren und den Items erfordern (Schnell et al., 2005, S. 162).

(Orchesterqualität, Chorqualität und die Qualität des Sologesangs) in Bezug auf die Ergebnisfaktoren (Klang und Tempo/ Rhythmus) aufeinander abzustimmen“ (Boerner, 2002, S. 93). Gleiches gilt für die Potentialfaktoren der szenischen Dimension, innerhalb derer die schauspielerische Qualität und die Ausstattungsqualität bspw. in Bezug auf die Handlung abzustimmen sind. Daher werden auf dieser Ebene zwei Faktoren erwartet, die untereinander vollkommen unkorreliert sind[1] und einen (möglichst) großen Teil der Gesamtstreuung der eingehenden Items erklären. Die Ergebnisse der Faktorenanalyse sind Tabelle 12 zu entnehmen.

Tabelle 12: Faktorenanalyse der zweiten Ebene der Oper „Die Zauberflöte“

Kennziffern der Items	Label der Items	fit2_Szene Faktorladungen	fit2_Musik Faktorladungen
fit2_2	Strahlkraft und Volumen	.12	**.81**
vfit2_3	Abstimmung Lautstärke	.26	**.83**
rfit2_6	übertönen	.12	**.73**
vfit2_8	Zudecken	.19	**.76**
vrfit2_9	Überflutung	.01	**.69**
fit3A1	Atmosphäre Bühnenbild	**.84**	.09
vfit2S_1	Handlung und Darsteller	**.89**	.01
fit2S_2	Dekoration	**.87**	.12
vfit3A2	Ausstattung	**.89**	.18
fit3A4	Abstimmung Kostüme - Bühnenbild	**.83**	.27
fit3A5	Requisiten und Licht	**.87**	.10
vrfit2S_4	Fremdkörper	**.79**	.17
vfit2S_3	stimmiges Bild	**.84**	.28
Varianzaufklärung		46%	24%

Anmerkungen: Extraktionsmethode: Hauptkomponentenanalyse. Rotationsmethode: Varimax mit Kaiser-Normalisierung. Vorgabe von zwei Faktoren. Die Formulierungen der Items können Anhang A entnommen werden.

Der Logik des analytischen Modells (Boerner, 2002) entsprechend lassen sich zwei Faktoren extrahieren: ‚Szene’ und ‚Musik’. Anders als bei Renz (2006, S. 47) (siehe Tabelle 13) bindet der Faktor ‚Szene’ einen erheblich größeren Teil der Varianz als der Faktor ‚Musik’ (46% vs. 24%; erklärte Gesamtvarianz: 70%). Im Vergleich zu ihrer Studie bestehen die Faktoren in der vorliegenden Untersuchung jeweils aus einer größeren Anzahl von Items, da keine Items

[1] Das an dieser Stelle verwendete Verfahren zur Faktorextraktion, die Hauptkomponentenanalyse, bildet lineare Kombinationen der Items. Dabei werden die Faktoren so ermittelt, dass sie untereinander vollkommen unkorreliert sind.

wegen der allzu häufig gewählten Antwortmöglichkeiten „habe ich nicht beachtet“ bzw. „kann ich nicht beurteilen“ ausgeschlossen werden müssen (vgl. Abschnitt 4.1.3).[1]

Die Hauptladungen sind – wie schon auf der ersten Ebene des Modells – hoch (≥ .69) und die Nebenladungen im Durchschnitt gering; die Items hängen also nur mit jeweils einem Konstrukt zusammen. Von der sog. Trennschärfe der Items kann daher auch von dieser Ebene her auf eine hohe Validität des modifizierten Fragebogens geschlossen werden.

Tabelle 13: Faktorenanalyse der zweiten Ebene der Oper „Aida“ – Renz (2006)

Kennziffern der Items	Label der Items	fit2_Szene Faktorladungen	fit2_Musik Faktorladungen
fit2_3	Abstimmung Lautstärke	-.02	**.91**
ufit2_7	„harmonisches Klangbild“	-.16	**.90**
fit2_2	Strahlkraft – Solisten	.04	**.69**
fit2S_1	Bühnenbild – Handlung	**.92**	-.06
fit2S_2	Dekoration / Requisiten Handlung	**.91**	-.01
Varianzaufklärung		34%	42%

Anmerkungen: Laienstichprobe: n = 39; Extraktionsmethode: Hauptkomponentenanalyse. Rotationsmethode: Varimax mit Kaiser-Normalisierung. Die Formulierungen der Items können Anhang A entnommen werden.
aus: Renz (2006, S. 47)

Faktorenanalyse der dritten Ebene des analytischen Modells

Auf der dritten Ebene des analytischen Modells (Boerner, 2002) befinden sich neben Konstrukten zur Qualität der Teilfaktoren einzelner Potential- und Ergebnisfaktoren auch Konstrukte zur Stimmigkeit innerhalb der jeweiligen Potentialfaktoren (siehe Abbildung 1). Insgesamt existieren auf dieser Ebene nach erfolgter Modifikation (vgl. Abschnitt 3.1.2) acht Konstrukte: ‚musikalische Leistung und Zusammenspiel des Orchesters sowie Leistung des Dirigenten’, ‚musikalische Leistung des Solisten’, ‚musikalische Leistung des Chores’, ‚szenische Leistung des Solisten’, ‚szenische Leistung des Chores’, ‚Szene_Ausstattung’ (Einzelaspekte), ‚Zusammenspiel der Solisten untereinander’ und ‚Zusammenspiel innerhalb des Chores’. Da sich jedoch bereits bei der Auswertung der deskriptiven Statistiken sowie bei der Analyse der Interkorrelation der Skalen zeigte, dass die Wahrnehmung des Publikums weniger differenziert ist als vom analytischen Modell angenommen, wird bei der

[1] So besteht bspw. der Faktor ‚Szene’ bei Renz (2006) nur aus zwei Items (siehe Tabelle 13), wohingegen in der vorliegenden Untersuchung sieben Items den Faktor ‚Szene’ bilden (siehe Tabelle 12).

Faktorenanalyse auf der dritten Ebene mit weniger als acht Faktoren gerechnet (siehe Tabelle 14).

Tabelle 14: Faktorenanalyse der dritten Ebene der Oper „Die Zauberflöte“

Kennziffern der Items	Label der Items	Faktor 1	Faktor 2	Faktor 3	Faktor 4	Faktor 5
mMO13	gute Vorbereitung	**.79**	.07	.27	.10	.09
vMO2	Lautstärke Orchester	**.75**	.40	.19	.05	.23
vuMO3	musikalische Spannung	**.89**	.23	.15	.14	.08
vMO5	Spannung und Intensität	**.84**	.28	.23	.15	.15
MO10	Lebendigkeit	**.85**	.31	.21	.12	.16
mfit3O11	Instrumentengruppen	**.84**	.30	.16	.09	.13
fit3O5	Orchester wie aus einem Guss	**.83**	.25	.24	.08	.19
MD1	im Griff	**.82**	.37	.15	.12	.04
vMD7	Balance	**.65**	.39	.31	.00	.37
MD13	dem Werk gewachsen	**.83**	.34	.20	.10	.15
vMS6	Unterschiede in der Lautstärke	.33	.41	**.71**	.19	.14
vuMS7	der Partie gewachsen	.16	.25	**.87**	.14	.01
MS10	Anspruch der Rolle	.25	.34	**.82**	.15	.10
MC2	zur Geltung	.36	**.83**	.23	.16	.11
mMC10	musikalische Leistung	.39	**.83**	.19	.20	.12
MC8	in jeder Hinsicht	.36	**.82**	.18	.23	.03
MC7	klangliche Brillanz	.36	**.79**	.24	.19	.05
SS1	eindringlich	.46	.44	**.47**	.41	.13
vSS4	Leben der Rolle	.29	**.48**	.28	.44	.19
SS5	Verkörperung	.50	.31	**.59**	.35	.06
vuSC1	Erscheinungsbild	.07	.09	.11	**.90**	.20
vSC2	voll abnehmen	.07	.19	.14	**.89**	.15
vSC3	ästhetisch befriedigend	.13	.20	.07	**.87**	.13
vSA13	Bühnenbild gefallen	.18	.03	.10	.40	**.76**
mSA16	Bühnenbild beeindruckend	.13	.08	.09	.35	**.85**
vSA15	Kostüme gefallen	.15	.24	.04	.49	**.59**
fit3C1	Homogenität	.38	**.61**	.32	.25	.27
vufit3C2	Tempo	.33	**.59**	.36	.07	.39
vfit3C3	Stimmen	.31	.26	.28	-.18	**.42**
vfit3S1	Abstimmung Solisten	**.58**	.11	.54	.02	.44
rfit3S3	Solisten gegeneinander	.38	.04	**.64**	.02	.29
ufit3S5	Harmonie	.35	.42	.36	-.13	**.45**
Varianzaufklärung		28%	17%	14%	12%	10%

Anmerkungen: Extraktionsmethode: Hauptkomponentenanalyse. Rotationsmethode: Varimax mit Kaiser-Normalisierung. Keine Vorgabe der Anzahl von Faktoren. Die Formulierungen der Items können Anhang A entnommen werden.

Wie erwartet liegt die Anzahl der gezogenen Faktoren deutlich unter acht; fünf Faktoren mit einem Eigenwert größer als 1 werden – nach Bereinigung von Doppel- und Nebenladungen[1] – durch die Hauptkomponentenanalyse extrahiert.[2] Insgesamt erklären sie eine Varianz von 81%, was deutlich über der erklärten Gesamtvarianz der Untersuchung von Renz (2006, S. 50)[3] und nur wenig unter der von Boerner et al. (in Druck)[4] liegt.

Inhaltlich lassen sich die Faktoren wie folgt interpretieren: Faktor 1, der am meisten Varianz bindet, umfasst Items zur Leistung und zum Zusammenspiel innerhalb des Orchesters sowie zur Leistung des Dirigenten. Die Zusammenlegung der im analytischen Modell Boerners (2002) getrennten Konstrukte ‚Orchester' und ‚Dirigent' wird bestätigt, nachdem eine Bestätigung dieser Vorgehensweise bereits durch die Reliabilitätsanalysen (vgl. Abschnitt 4.2.1) und die Analyse der Interkorrelation der Skalen (vgl. Abschnitt 4.2.2) erfolgte. Zu einer vom Orchester unabhängigen Betrachtung der Leistung des Dirigenten ist die Mehrheit des Publikums also nicht in der Lage, genauso wenig wie zur von der musikalischen Leistung des Orchesters differenzierten Betrachtung seines Zusammenspiels. Vermutet wird die Ursache dafür in dazu benötigtem Wissen und nötiger Erfahrung, die der Mehrheit des Publikums nach eigener Einschätzung fehlen. Ob die Teilstichprobe der Experten, denen das notwenige Wissen bzw. die Erfahrung zugeschrieben werden können, zu einer differenzierteren Wahrnehmung in der Lage ist, wie von der Betrachtung der Häufigkeiten der Antwortmöglichkeiten „habe ich nicht beachtet" bzw. „kann ich nicht beurteilen" angedeutet, wird im weiteren Verlauf zu klären sein.

Der zweite Faktor wird von Items zur musikalischen Leistung sowie zum Zusammenspiel innerhalb des Chores gebildet und erklärt eine Varianz von 17%. Auch der Chor wird demnach nicht in so differenzierter Weise betrachtet wie theoretisch angenommen (Boerner,

[1] Es gingen nicht sämtliche Items in die Faktorenanalyse ein, die nach der Reliabilitätsanalyse die entsprechenden Skalen konstituierten (siehe Tabelle 9). Auf vier Items musste wegen hoher Doppel- und Nebenladungen verzichtet werden (MS1, MS5, vSS8 und mSA17). Im Rahmen erneuter Reliabilitätsanalysen wurde jedoch sichergestellt, dass sich Cronbachs Alpha aufgrund des Ausschlusses dieser Items jeweils nur unwesentlich verschlechterte. Außerdem gilt nach wie vor das Kriterium, dass sich eine Skala aus mindestens drei Items zusammensetzen muss.

[2] Faktorenwerte mit einem Eigenwert unter 1 erklären einen geringeren Betrag der gesamten Streuung als die einzelnen Items, die zumindest sich selbst erklären, so dass nur Faktoren mit einem Eigenwert größer 1 sinnvoll zu bilden sind. Der Grund hierfür ist, dass Variablenwerte bei einer Faktorenanalyse in Z-Werte transformiert werden und eine einzelne Variable daher immer genau eine Streuung von 1 erklärt.

[3] Renz (2006) extrahiert vier Faktoren, die eine Varianz von 76% aufklären. Dieses sind die Faktoren ‚Szene_Ausstattung', ‚Chor – Szene', ‚Solisten' und ‚Orchester'.

[4] In jener Studie können sechs Faktoren (‚Solist – Musik', ‚Chor – Szene', ‚Orchester', ‚Solist – Szene', ‚Zusammenspiel innerhalb des Chores' und ‚Zusammenspiel der Solisten untereinander') gezogen werden, die eine Gesamtvarianz von 82% erklären.

2002). Ähnliches deutete bereits die Analyse der Interkorrelation der Skalen an, bei der die Skalen ‚musikalische Leistung des Chores' und ‚Zusammenspiel innerhalb des Chores' stark korrelieren ($r = .67$, $p \leq 0.01$). In noch höherem Ausmaß gilt dies für die Leistung des Solisten (Faktor 3). Hierbei bilden dessen musikalische und szenische Leistung sowie das Zusammenspiel aller Solisten einen Faktor. Einzige Ausnahmen bilden die Items ‚Papageno lebte die Rolle so, dass ich mich sehr gut mit ihr identifizieren konnte' (vSS4) und ‚Das Solistenensemble wirkte musikalisch untereinander ausgesprochen harmonisch' (ufit3S5), die auf den Faktoren 2 und 4 bzw. 2 und 5 laden und bei einer weiteren Modifikation des Fragebogens überarbeitet werden müssten.

Items zur szenischen Leistung des Chores bilden Faktor 4, so dass zumindest bei dem Akteur ‚Chor' von einer getrennten Betrachtung der Dimensionen ‚Musik' und ‚Szene' gemäß dem analytischen Modell (Boerner, 2002) ausgegangen werden kann. Faktor 5 konstituiert sich aus den Items zur Szene bzw. Ausstattung, die unabhängig von der Stimmigkeit, d.h. jeweils für sich genommen, einen ästhetischen Wert haben, und klärt 10% der Varianz auf.

Neben der Mehrzahl der Items, die – abgesehen von geringen Nebenladungen – auf exakt einem Faktor laden, fallen mehrere Items auf, die dies nicht tun. Es sind zusätzlich zu den bereits erwähnten Items insbesondere Item vMS6, welches (zusätzlich zum erwarteten Faktor ‚Solist') auch auf Faktor 2 (musikalische Leistung und Zusammenspiel innerhalb des Chores) lädt, Item SS1 (zusätzlich zum Faktor ‚Solist' auch auf den Faktoren 1, 2 und 4), Item SS5 (neben Faktor ‚Solist' auch Faktor 1), Item vSA15 (zusätzlich zur Hauptladung auf Faktor ‚Szene_Ausstattung' hohe Nebenladung auf Faktor 4) sowie Item vfit3S1 (neben Faktor ‚Solist' auch Faktor 1). Darüber hinaus fallen drei Items ins Auge, die auf anderen als den erwarteten Faktoren laden.[1]

Abgesehen von diesen Items, die bei einer erneuten Modifikation des Fragebogens besondere Aufmerksamkeit erforderten, spricht die Trennschärfe der Mehrheit der Items auch auf der dritten Ebene für eine hohe Validität des modifizierten Fragebogens. Vergleichbar zu den Untersuchungen von Renz (2006) und Boerner et al. (in Druck) werden die physisch unterscheidbaren Konstrukte ‚Orchester', ‚Chor', ‚Solisten' und ‚Ausstattung' vom Publikum unterschieden, wobei beim Chor darüber hinaus dessen unterschiedliche Funktionen differenziert wahrgenommen werden. Das Orchester, zusammengefasst mit dem Dirigenten,

[1] Dieses sind die Items vSS4, vfit3C3 und ufit3S5.

bildet den Faktor mit der größten Erklärungskraft, obwohl gerade dessen Bewertung insbesondere den Laien (vgl. Abschnitt 4.1.3) Schwierigkeiten bereitete.

Faktorenanalysen der Konstrukte zur Erfassung des Gesamturteils des Publikums über einen Abend in der Oper (‚Gesamterlebnis Opernabend')

Nach Bestätigung der Validität der Konstrukte zur Erfassung der künstlerischen Qualität einer Aufführung im Musiktheater werden nun Faktorenanalysen bei den Konstrukten zur Erfassung des ‚Gesamterlebnisses Opernabend' durchgeführt. Aus Platzgründen sind deren Ergebnisse in Anhang H dargestellt. Wie erwartet lässt sich bei den Konstrukten ‚Gesamturteil', ‚künstlerische Qualität', ‚Komplexität', ‚Neuartigkeit' und ‚Identifikation' ohne Vorgabe der Anzahl von Faktoren, d.h. explorativ, jeweils nur ein Faktor extrahieren, der zwischen 66% (Konstrukt ‚Komplexität') und 86% (‚Gesamturteil') der Gesamtvarianz erklären kann (siehe Anhang H, Tabellen 32 - 37). Zu einem anderen Ergebnis kommt die Faktorenanalyse beim Konstrukt ‚situative Rahmenbedingungen', wo zwei Faktoren gezogen werden. Diese lassen sich inhaltlich interpretieren als Faktor ‚Aufführungssaal' mit Items u.a. zur Akustik des Saales und zum Komfort des Sitzes sowie Faktor ‚Umfeld' (Service des Hauses, Räumlichkeiten, Atmosphäre) (siehe Anhang H, Tabelle 38) und gehen getrennt in die sich anschließenden Pfadanalysen ein. Die Erwartung, die zur emotionalen Wirkung der Aufführung erhobenen Items würden in die Faktoren ‚Empathie', ‚persönliche Assoziationen', ‚Genussfaktor', ‚Ich-Beteiligung', ‚vorübergehende Gefühlsregung' und ‚Gefühl des Bewegt-Seins' zerfallen, bestätigt sich nicht. Stattdessen können drei Faktoren extrahiert werden, die gemeinsam 73% der Gesamtvarianz erklären (siehe Anhang H, Tabelle 34). Von diesen können zwei inhaltlich sinnvoll interpretiert werden: Faktor ‚Genuss/ Ich-Beteiligung' und Faktor ‚Empathie/ persönliche Assoziationen', die jeweils als unabhängige Variablen in die später durchgeführten Pfadanalysen eingehen.

Zusammenfassend lässt sich eine hohe Validität des weiterentwickelten Fragebogens konstatieren, zu deren Überprüfung die Faktorenanalysen durchgeführt wurden. Um die weiteren Untersuchungsfragen inhaltlich zu beantworten, müssten im Folgenden eine erneute Weiterentwicklung des Fragebogens sowie eine weitere Datenerhebung erfolgen. Da dies im Rahmen der vorliegenden Arbeit nicht möglich ist und zudem die Struktur des analytischen Modell Boerners (2002) nach den bisher erfolgten Analysen bestätigt werden kann, werden die weiteren Untersuchungsfragen anhand der bereits erhobenen Daten zu klären versucht (Kapitel 5).

5. Analysen zur Klärung der Untersuchungsfragen

5.1 Ergebnisse zu Untersuchungsfrage 1: Gegenstände der Bewertung künstlerischer Qualität in der Oper

Nach der Präsentation der Ergebnisse zu Untersuchungsfrage 1a – welche musikalischen und szenischen Aspekte einer Opernaufführung die Rezipienten bei ihren Urteilen in welchem Differenzierungsgrad beachten und unterscheiden – in Abschnitt 5.1.1 wird in Abschnitt 5.1.2 Untersuchungsfrage 1b – mit welchem Gewicht unterschiedliche Teilaspekte in das Urteil über die künstlerische Qualität eingehen – beantwortet werden. Wie bereits Moser (2005, S. 57f.) und Renz (2006, S. 59) für ihr Stimulusmaterial gezeigt haben, wird auch für die bewertete Aufführung der Oper „Die Zauberflöte“ vermutet, dass die verschiedenen Teilaspekte vom Publikum nicht gleich gewichtet werden, sondern unterschiedlich wichtig für die Bewertung der Gesamtqualität sind.

5.1.1 Untersuchungsfrage 1a: musikalische und szenische Aspekte einer Opernaufführung

Erkenntnisse zu Untersuchungsfrage 1a lassen sich sowohl aus der deskriptiven Auswertung (vgl. Abschnitt 4.1.3) als auch aus der Interkorrelation der Skalen (vgl. Abschnitt 4.2.2) und den Faktorenanalysen (vgl. Abschnitt 4.2.3) gewinnen. So zeigte die Analyse der Häufigkeit des Auftretens der Antwortmöglichkeiten „kann ich nicht beurteilen“, dass die Bewertung der musikalischen Aspekte einer Opernaufführung sowohl den Laien als auch den Experten deutlich mehr Schwierigkeiten bereitet als die Bewertung der szenischen Aspekte. Insgesamt lässt sich schließen, dass zur Beurteilung der künstlerischen Qualität einer Opernaufführung eine Vielzahl von Konstrukten herangezogen wird, die durch das analytische Modell Boerners (2002) bzw. den weiterentwickelten Fragebogen sehr gut abgebildet werden.

Genauere Erkenntnisse zu Untersuchungsfrage 1a ergaben sich aus den Faktorenanalysen. So zeigte sich, dass die Befragten auf der ersten Ebene des Modells (Boerner, 2002) die ‚Stimmigkeit zwischen der musikalischen und der szenischen Funktion’ (fit1) und auf der zweiten Ebene die ‚Stimmigkeit innerhalb der musikalischen Dimension’ (fit2_M) sowie die ‚Stimmigkeit innerhalb der szenischen Dimension’ (fit2_S) unterscheiden. Darüber hinaus ergab sich für die dritte Ebene des Modells, dass das Publikum die physisch unterscheidbaren Konstrukte Orchester/ Dirigent, Chor, Solist und Ausstattung getrennt wahrnimmt sowie beim Chor zwischen dessen unterschiedlichen Funktionen (musikalische Leistung/Zusammenspiel innerhalb des Chores, szenische Leistung) differenziert. Zusammenfassend kann festgehalten

werden, dass sowohl auf der ersten als auch auf der zweiten Ebene die Differenziertheit den theoretischen Erwartungen des analytischen Modells (Boerner, 2002; vgl. Abschnitt 2.2.1) entspricht und lediglich die Konstrukte der dritten Ebene nicht in erwartet differenzierter Weise vom Publikum wahrgenommen werden.

5.1.2 Untersuchungsfrage 1b: Gewichtung der Teilaspekte

Ausgehend von den in Abschnitt 4.2.3 durchgeführten Faktorenanalysen werden nun Skalen gebildet, deren Einfluss auf das Urteil über die künstlerische Qualität überprüft werden wird (Untersuchungsfrage 1b).[1] Welche Items bei der Skalenbildung verwendet wurden und die jeweilige Höhe von Cronbachs Alpha kann Tabelle 15 entnommen werden.

Tabelle 15: Skalenbildung der Faktoren

Skala	Cronbachs Alpha	Kennziffern der Items der jeweiligen Skala
Gesamtqualität	.90	mGQ1, mGQ2, vNG2
fit1	.92	vfit1_1, fit1_3, fit1_6, fit1_7, vfit1_8
fit2_Szene	.95	fit3A1, vfit3A2, fit3A4, fit3A5, vfit2S_1, fit2S_2, vrfit2S_4, vfit2S_3
fit2_Musik	.82	fit2_2, vfit2_3, rfit2_6, vfit2_8, vrfit2_9
Orchester	.97	mMO13, vMO2, vuMO3, vMO5, MO10, mfit3O11, fit3O5, MD1, vMD7, MD13
Chor (Musik, fit3)	.93	MC2, mMC10, MC8, MC7, fit3C1, vufit3C2, vfit3C3
Solist	.90	vMS6, vuMS7, MS10, SS1, vSS4, SS5, vfit3S1, rfit3S3, ufit3S5
Chor (Szene)	.94	vuSC1, vSC2, vSC3
Szene – Ausstattung	.88	vSA13, mSA16, vSA15

Anmerkung: fit 1 = Stimmigkeit zwischen der musikalischen und der szenischen Dimension, fit2_Szene = Stimmigkeit innerhalb der szenischen Dimension, fit2_Musik = Stimmigkeit innerhalb der musikalischen Dimension. Die Formulierungen der Items sind in Anhang A zu finden.

Bei Anwendung des häufig verwendeten Kriteriums, nach dem zusammengesetzte Skalen ab einem Wert von $\alpha = 0.8$ als hinreichend zuverlässig angesehen werden, können sämtliche Skalen als hoch reliabel bezeichnet werden. Dies ist umso bemerkenswerter, als dass sich ein Teil der Skalen lediglich aus drei Items zusammensetzt (vgl. Abschnitt 4.2.1).

[1] Die an dieser Stelle durchgeführte Skalenbildung unterscheidet sich von jener aus Abschnitt 4.2.1 insofern, als dass nun nicht auf Grundlage der im analytischen Modell Boerners (2002) theoretisch angenommenen Konstrukte, sondern auf Basis der erfolgten Faktorenanalysen Skalen gebildet werden.

Vor der Durchführung von Analysen zur Erhebung der Gewichtung der Teilaspekte werden zunächst die Beziehungen zwischen den Skalen überprüft, da von diesen die Wahl des geeigneten Analyseinstrumentes abhängt (siehe unten). Insbesondere ist dabei von Interesse, ob die Struktur des Publikumsurteils hierarchisch ist, wie vom analytischen Modell Boerners (2002) angenommen, oder nicht.[1] Tabelle 16 gibt eine Übersicht über die Korrelationen zwischen den Skalen.

Tabelle 16: Korrelationsmatrix der Skalen

	fit1	fit2_S	fit2_M	Orchester	Chor (Musik, fit3)	Solist	Chor (Szene)	Ausstattung
fit1	1							
fit2_S	.80**	1						
fit2_M	.52**	.35**	1					
Orchester	.50**	.32**	.64**	1				
Chor (Musik, fit3)	.53**	.42**	.53**	.76**	1			
Solist	.68**	.50**	.61**	.78**	.74**	1		
Chor (Szene)	.58**	.56**	.25**	.25**	.43**	.41**	1	
Ausstattung	.71**	.83**	.25**	.39**	.43**	.49**	.44**	1

Anmerkung: **$p \leq 0.01$. fit 1 = Stimmigkeit zwischen der musikalischen und der szenischen Dimension, fit2_S = Stimmigkeit innerhalb der szenischen Dimension, fit2_M = Stimmigkeit innerhalb der musikalischen Dimension. Die Bildung der Faktoren kann Abschnitt 4.2.3 entnommen werden.

Gemäß der Logik des analytischen Modell Boerners (2002), wonach die Teilaspekte der dritten Ebene jeweils entweder auf die Stimmigkeit innerhalb der musikalischen oder innerhalb der szenischen Dimension (Ebene 2) einwirken, die ihrerseits die Stimmigkeit zwischen Szene und Musik (Ebene 1) beeinflussen (siehe Abbildung 1), korrelieren die Skalen der ersten und zweiten Ebene sowohl untereinander als auch mit den Skalen der dritten Ebene zum Großteil hoch, was als erster Hinweis auf eine hierarchische Struktur des Publikumsurteils über die künstlerische Qualität einer Opernaufführung interpretiert werden kann. Anders als erwartet korrelieren jedoch ebenfalls die Skalen der dritten Ebene untereinander teilweise hoch. So korreliert die Skala ‚Orchester' sowohl mit der Skala ‚Chor', die sich aus dessen musikalischer Leistung und dem Zusammenspiel innerhalb des Chores zusammensetzt ($r = .76$, $p \leq 0.01$), als auch mit der Skala ‚Solist' ($r = .78$, $p \leq 0.01$). Auch

[1] Wie in Abschnitt 2.1.1 erläutert, stellt in diesem jeder untergeordnete fit eine notwendige, jedoch nicht hinreichende Bedingung für den fit auf der nächst höheren Ebene dar (Boerner, 2002, S. 97).

zwischen den Skalen ‚Chor' (musikalische Leistung, Zusammenspiel) und ‚Solist' besteht eine hohe Korrelation ($r = .74$, $p \leq 0.01$). Da die Wahl eines geeigneten Analyseinstrumentes zur Messung der Gewichtung der Teilaspekte davon abhängt, ob die Struktur des Publikumsurteils hierarchisch ist oder nicht[1] und sich sowohl aus dem theoretischen Modell (Boerner, 2002) als auch aus der Korrelationsmatrix Hinweise auf eine Hierarchie ergeben, wird zu deren weiterer Überprüfung im Folgenden eine Pfadanalyse durchgeführt.

Pfadanalysen

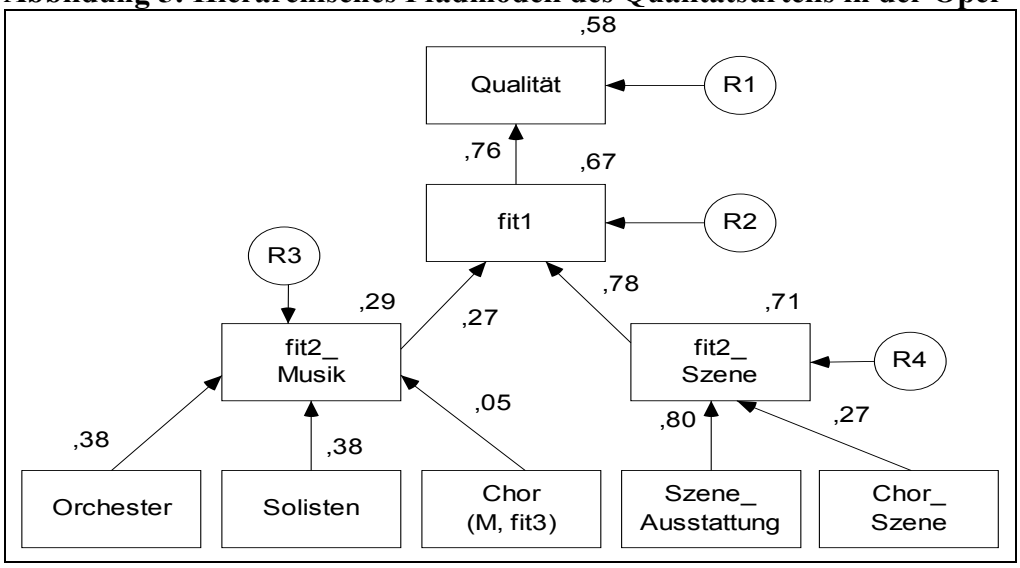
Anders als Regressionsanalysen sind Pfadanalysen in der Lage, nicht nur Aussagen über Beziehungen zwischen einer abhängigen und einer bzw. mehreren unabhängigen Variablen zu treffen, sondern zusätzlich zu testen, wie verschiedene abhängige Variablen wieder auf andere Variablen einwirken (Schnell et al., 2005, S. 461). Dazu werden in einem ersten Schritt die aus theoretischen Überlegungen hergeleiteten Kausalhypothesen in einem sog. Pfaddiagramm zusammengefasst. In einem zweiten Schritt wird anschließend getestet, ob sich das Modell durch die erhobenen Daten bestätigen lässt (Bortz, 1999, S. 456). Im Gegensatz zu struktur-entdeckenden Verfahren, bei denen im Voraus keinerlei Vorstellungen über Zusammenhänge zwischen Variablen bestehen, arbeiten Pfadanalysen somit struktur-prüfend bzw. hypothesen-testend, d.h. konfirmatorisch. Anhand von Korrelationen zwischen Variablen sollen dabei Aussagen über Kausalitäten getroffen werden, weshalb Pfadanalysen auch als ‚Kausalmodelle' bezeichnet werden.[2]

Abbildung 5 zeigt das hierarchische Pfadmodell des Publikumsurteils über künstlerische Qualität in der Oper, das auf Grundlage des analytischen Modell Boerners (2002) (vgl. Abschnitt 2.2.1) sowie der in Abschnitt 4.2.3 erfolgten Faktorenanalysen erstellt wurde.

[1] Ursächlich dafür sind die starken Korrelationen der Konstrukte, die mit einer hierarchischen Struktur des Publikumsurteils einhergingen. Diese würden bspw. die Durchführung einer linearen Regressions-analyse ausschließen, da bei dieser die Schätzungen der Regressionsparameter mit zunehmender Kollinearität unzuverlässiger werden (Backhaus et al., 2003, S. 88).

[2] Da aus einem statistischen Modell über Korrelationen streng genommen jedoch niemals Aussagen über Kausalbeziehungen getroffen werden können (vgl. Abschnitt 4.2.2), ist diese Bezeichnung – obwohl gebräuchlich – falsch (Schnell et al., 2005, S. 463). Weil aber Korrelationen zwischen Variablen eine notwendige, wenn auch nicht hinreichende Bedingung für Kausalität darstellen, dienen sie als Grundlage von Pfadanalysen, wenngleich sich Kausalität damit niemals bestätigen, sondern ausschließlich widerlegen lässt. Da sich in der Regel jedoch mehrere Kausalmodelle finden lassen, die mit der von einer Pfadanalyse ermittelten Kovarianz- und Korrelationsstruktur in Einklang zu bringen sind, werden in der vorliegenden Arbeit Korrelationen lediglich als ‚Hinweise' für Kausalitäten interpretiert.

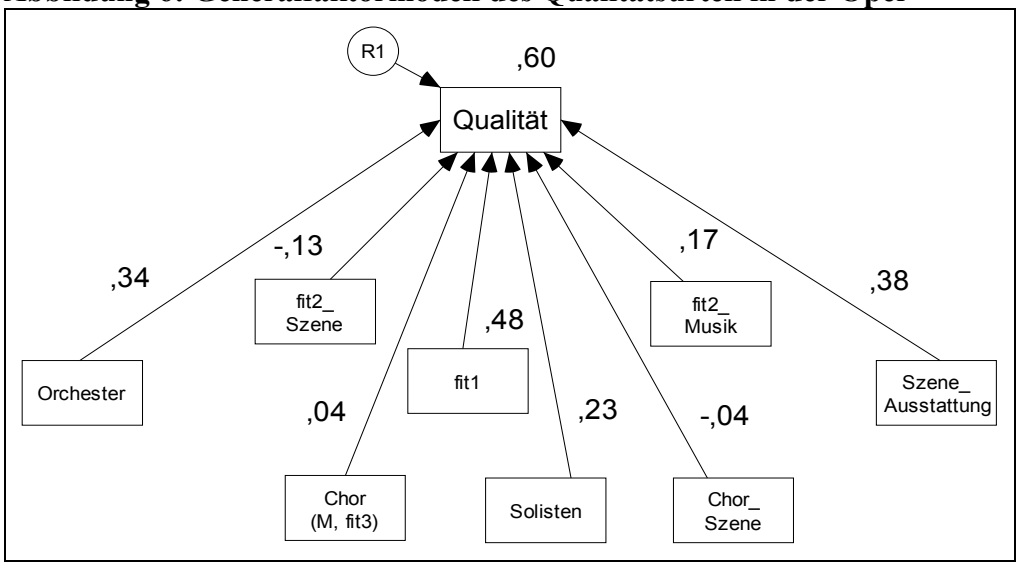
Abbildung 5: Hierarchisches Pfadmodell des Qualitätsurteils in der Oper



Anmerkungen: Rechtecke präsentieren manifeste Variablen, Pfeile stellen ihre Zusammenhänge dar. Werte bei den Pfeilen stellen die standardisierten ß-Koeffizienten dar. Werte am rechten oberen Rand der manifesten Variablen repräsentieren die Kommunalität. ‚R' sind Residualvariablen.

Zur Überprüfung, ob die Konstrukte hierarchisch auf das Publikumsurteil wirken oder nicht, erscheint ein Vergleich des hier gebildeten Modells mit einem sog. Generalfaktormodell, in welchem sämtliche Konstrukte direkten Einfluss auf das Urteil über die künstlerische Qualität nehmen und demnach keinerlei hierarchische Beziehungen bestehen, sinnvoll.[1] Abbildung 6 gibt das Generalfaktormodell grafisch wieder.

Abbildung 6: Generalfaktormodell des Qualitätsurteil in der Oper



Anmerkungen: Rechtecke präsentieren manifeste Variablen, Pfeile stellen ihre Zusammenhänge dar. Werte bei den Pfeilen stellen die standardisierten ß-Koeffizienten dar. Werte am rechten oberen Rand der manifesten Variablen repräsentieren die Kommunalität. ‚R' sind Residualvariablen.

[1] Da an dieser Stelle speziell von Interesse ist, ob die Konstrukte – wie vom analytischen Modell Boerners (2002) angenommen – hierarchisch auf das Gesamturteil wirken oder nicht, sind die ß-Koeffizienten zwar in Abbildung 5 enthalten, werden an dieser Stelle jedoch nicht interpretiert.

Die Beurteilung der Übereinstimmung der beiden Modelle mit den empirisch erhobenen Daten (sog. Anpassungsgüte) lässt sich mit Gütemaßen durchführen, wobei im Gegensatz zu vergleichbaren Analysen bei Pfadanalysen eine Vielzahl von Gütemaßen existiert, die zu durchaus unterschiedlichen Ergebnissen kommen können (Schermelleh-Engel et al., 2003, S. 23). Eine Auswahl geeigneter Kriterien kann Tabelle 17 entnommen werden.[1]

Tabelle 17: Anpassungsgüte des Modellvergleichs

	χ^2	df	p	χ^2/df	RMSEA	NFI	CFI	AIC	ECVI
hierarchisches Modell (Abbildung 5)	320.07	28	.000	11.43	.296	.582	.595	372.07	3.13
Generalfaktormodell (Abbildung 6)	612.81	28	.000	21.89	.419	.200	.189	664.81	5.59

Anmerkungen: χ^2 = Chi-Quadrat-Wert, df = Freiheitsgrade, p = Wahrscheinlichkeit, RMSEA = Root Mean Square Error of Approximation, NFI = Normed Fit Index, CFI = Comparative Fit Index, AIC = Akaike Information Criterion, ECVI = Expected Cross Validation Index.

Das einzige Gütekriterium mit integriertem Signifikanztest ist der Chi-Quadrat-Test, der die Nullhypothese prüft, nach der die empirische Kovarianzmatrix der modelltheoretischen entspricht. Chi-Quadrat-Werte (χ^2) von 320.07 bzw. 612.81 wurden errechnet; die Nullhypothese muss demnach in beiden Fällen abgelehnt werden (p = .000). Da die Ergebnisse des Chi-Quadrat-Tests generell jedoch erheblich von der Stichprobengröße, der Parameteranzahl sowie der Verteilung der Daten abhängen, müssen ebenfalls die Verhältnisse zwischen χ^2 und Freiheitsgraden (jeweils df = 28) in Betracht gezogen werden. Im vorliegenden Fall liegen diese bei χ^2/df = 11.43 bzw. 21.89. Die Tatsache, dass kleinere Werte für eine bessere Anpassungsgüte des Modells mit den empirischen Daten stehen (Homburg & Baumgartner, 1995, S. 170), macht die Überlegenheit des hierarchischen Modells gegenüber dem Generalfaktormodell deutlich.

Anders als der Chi-Quadrat-Test, der zu den ‚absoluten' Gütekriterien zählt, setzt der Root Mean Square Error of Approximation (RMSEA) keine *exakte* Übereinstimmung zwischen theoretischem Modell und empirischen Daten voraus, die in den Sozialwissenschaften nur in seltenen Fällen anzutreffen sein wird, sondern testet deren *ungefähre* Übereinstimmung. Dass auch hier kleinere Werte eine bessere Anpassungsgüte des Modells indizieren (Browne & Cudeck, 1993), bestätigt die Überlegenheit des hierarchischen Modells.

[1] Ob sich ein Gütemaß für die Messung der Übereinstimmung zwischen Modell und empirischen Daten eignet, hängt insbesondere von der Wahl des Schätzalgorithmus ab, der die Differenz zwischen der mit Hilfe der geschätzten Parameter berechenbaren modelltheoretischen Korrelationsmatrix und der empirischen Korrelationsmatrix zu minimieren versucht (Backhaus et al., 2003, S. 370). In dieser Arbeit wurde die Maximum-Likelihood-Methode gewählt, die das gebräuchlichste iterative Schätzverfahren darstellt.

Zum selben Ergebnis kommen auch die inkrementellen Gütekriterien NFI (Normed Fit Index) und CFI (Comparative Fit Index), die die Übereinstimmung zwischen Theorie und Empirie im Vergleich zu Modellen mit weniger Parametern, d.h. Basismodellen, messen, sowie die speziell zum Vergleich mehrerer Modelle entwickelten Gütekriterien AIC (Akaike Information Criterion) und ECVI (Expected Cross Validation Index) (Schermelleh-Engel et al., 2003, S. 43f.). Im Gegensatz zu den übrigen Kriterien sind bei beiden letztgenannten unstandardisiert; niedrigere Werte stehen für das jeweils geeignetere, d.h. die empirischen Daten besser abbildende, Modell.

Die deutliche Überlegenheit des hierarchischen Pfadmodells lässt auf eine hierarchische Struktur des Publikumsurteils in der Oper schließen. Demnach stellen zur Erfassung der Gewichtung der Teilaspekte für das Urteil des Publikums über die künstlerische Qualität einer Opernaufführung Strukturgleichungsmodelle das passende Analyseinstrument dar, da bei diesen – im Gegensatz zu linearen Regressionsanalysen – die Schätzungen der Parameter mit zunehmender Kollinearität nicht unzuverlässiger werden.

Strukturgleichungsmodelle

Strukturgleichungsmodelle kombinieren regressionsanalytische Überlegungen mit Aspekten der Faktoren- und Pfadanalyse und stellen daher eine geeignete Methode zur Überprüfung der Übereinstimmung eines theoretisch entwickelten Modells mit empirisch erhobenem Datenmaterial dar (Bortz, 1999, S. 456). Im Gegensatz zu Pfadanalysen bieten sie jedoch zusätzlich die Möglichkeit, neben Beziehungen zwischen manifesten Variablen auch solche zwischen latenten, d.h. nicht direkt messbaren, Variablen zu überprüfen, wie im vorliegenden Fall erwünscht (Bortz, 1999, S. 456). Strukturgleichungsmodelle bestehen daher aus zwei Teilen: Als *Messmodell* wird der Teil bezeichnet, der die manifesten Variablen den ihnen zu Grunde liegenden Konstrukten, d.h. latenten Variablen, zuordnet. Diese sind damit den Faktoren im Sinne der Faktorenanalyse vergleichbar; zusätzlich können resultierende Pfadkoeffizienten als Faktorladungen, d.h. als Ergebnisse einer konfirmatorischen Faktorenanalyse, interpretiert werden. Die Ergebnisse der in Abschnitt 4.2.3 durchgeführten explorativen, d.h. ohne exakte Hypothesen über das zu erwartende Ergebnis vollzogenen, Faktorenanalyse können daher mit Hilfe eines Strukturgleichungsmodells auf ihre Korrektheit überprüft werden. Den zweiten Teil eines Strukturgleichungsmodells bildet das sog. *Strukturmodell*, welches die latenten Variablen zueinander in Beziehung setzt und mit regressionsanalytischen Verfahren verglichen werden kann. Mit dessen Hilfe wird geklärt

werden, mit welcher Gewichtung die unterschiedlichen Teilaspekte in das Gesamturteil der Rezipienten über die künstlerische Qualität einer Opernaufführung eingehen (Untersuchungsfrage 1b).

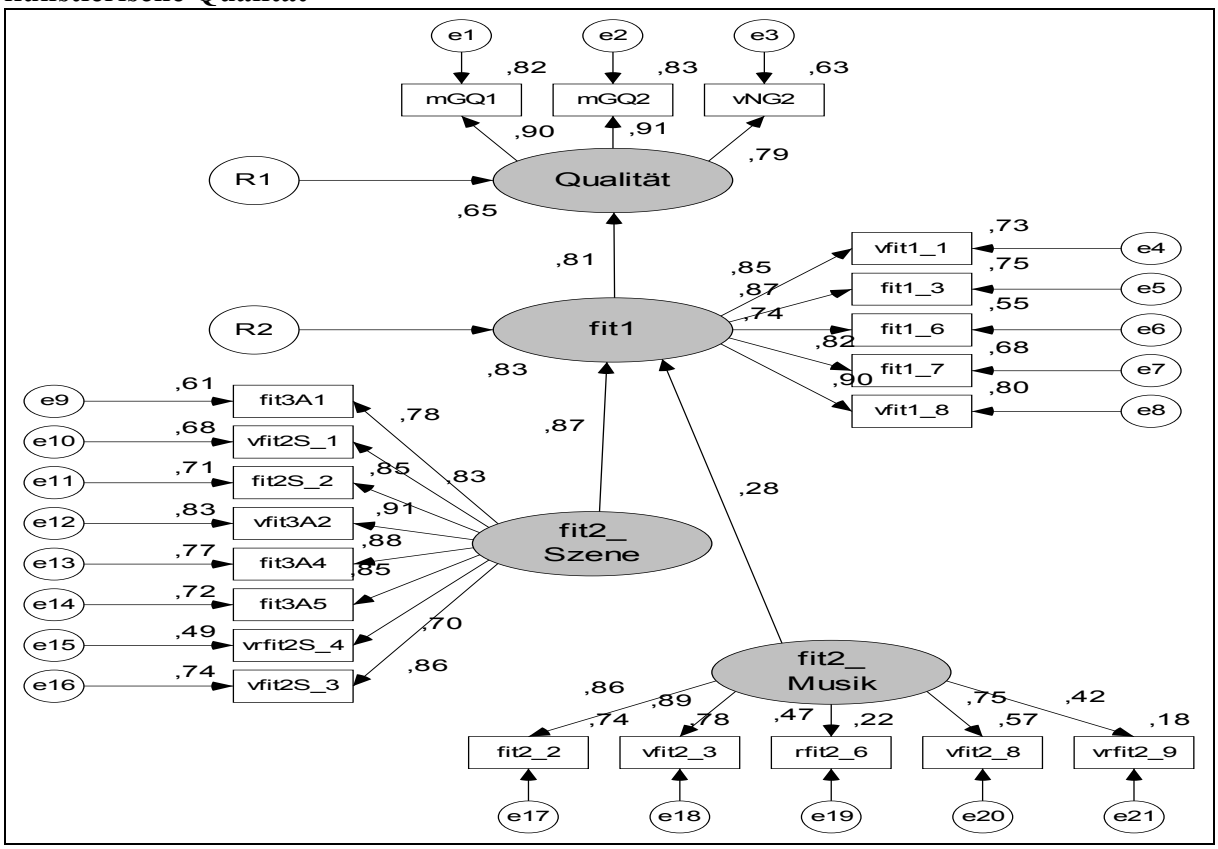
Zur korrekten Berechnung der Zusammenhänge zwischen Variablen eines Strukturgleichungsmodells muss neben einer Normalverteilung der Daten und der Absenz systematisch fehlender Werte auch eine ausreichend große Stichprobe gegeben sein. Häufig wird in der Literatur zur Parameterschätzung ein Stichprobenumfang von $n \geq 100$, gelegentlich auch $n \geq 200$, und/oder fünf Befragte pro zu schätzendem Parameter gefordert (Backhaus et al., 2003, S. 364), so dass die Verwendung eines Strukturgleichungsmodells aufgrund der hohen Anzahl zu schätzender Parameter im Rahmen dieser Arbeit nicht möglich ist.[1] Mit welcher Gewichtung die unterschiedlichen Teilaspekte in das Gesamturteil der Rezipienten über die künstlerische Qualität einer Opernaufführung eingehen, bleibt somit vorläufig unklar.

Wenngleich also die Gesamtstruktur des Modells anhand von Strukturgleichungsmodellen nicht ermittelt werden kann, ist dies doch für Teilstrukturen des Gesamtmodells aufgrund der geringeren Anzahl zu schätzender Parameter möglich. So können der Einfluss der zweiten auf die erste Ebene sowie der Einfluss der ersten Ebene auf das Urteil über die künstlerische Qualität quantifiziert werden (siehe Abbildung 7)[2], wobei die latenten Variablen ‚Qualität', ‚fit1', ‚fit2_Szene' und ‚fit2_Musik' jeweils durch die manifesten Items gemessen werden, die den Faktoren im Rahmen der in Abschnitt 4.2.3 erfolgten explorativen Faktorenanalyse zugeordnet wurden.

[1] Bei fünf Befragten pro zu schätzendem Parameter wäre zur korrekten Berechung eines Strukturgleichungsmodells ein Stichprobenumfang von n = 265 notwendig, der in der vorliegenden Arbeit bei weitem nicht erreicht wird (n = 120, vgl. Abschnitt 4.1.1).

[2] Ursächlich ist die geringere Anzahl zu schätzender Parameter, die bei dieser Teilstruktur einen Stichprobenumfang von n = 105 notwendig macht, vorausgesetzt es werden fünf Befragte pro zu schätzendem Parameter gefordert (Backhaus et al., 2003, S. 364).

Abbildung 7: Teilstruktur I des Strukturgleichungsmodells des Urteils über die künstlerische Qualität



Anmerkungen: Grau schraffierte Ellipsen präsentieren latente Variablen, Rechtecke präsentieren manifeste Variablen, ‚e' stellen den jeweiligen Messfehler dar, ‚R' sind Residualvariablen. Werte am linken unteren Rand der latenten Variablen stellen die erklärte Varianz dar, Werte am rechten oberen bzw. am linken Rand der manifesten Variablen repräsentieren die Kommunalität. Pfeile stellen Pfade dar, Werte bei den Pfeilen des Strukturmodells stellen die standardisierten ß-Koeffizienten dar, Werte bei den Pfeilen der Messmodelle repräsentieren die Faktorladungen. Es wurden die Kennziffern der jeweiligen Items verwendet; die Formulierungen können Anhang A entnommen werden.

In der hier abgebildeten standardisierten Lösung geben die Regressionskoeffizienten im *Messmodell* Auskunft über die Stärke der Korrelationen der manifesten Variablen mit den hypothetisch angenommenen Konstrukten und können wie oben erläutert wie Faktorladungen interpretiert werden. Da sämtliche Werte – zum überwiegenden Teil deutlich – über .40 liegen, können die Ergebnisse der explorativen Faktorenanalyse bestätigt werden; die Befragten unterscheiden die ‚Stimmigkeit zwischen der musikalischen und der szenischen Funktion' (fit1) auf der ersten Ebene des Modells (Boerner, 2002) und die ‚Stimmigkeit innerhalb der musikalischen Dimension' (fit2_M) sowie die ‚Stimmigkeit innerhalb der szenischen Dimension' (fit2_S) auf der zweiten Ebene. Eine umfassende Modellvereinfachung der Teilstruktur vergleichbar zu Renz (2006, S. 63) erscheint demnach nicht notwendig.

Zur Klärung der Untersuchungsfrage 1b können in einem zweiten Schritt die Koeffizienten des *Strukturmodells* betrachtet werden: So erklären die Konstrukte ‚Stimmigkeit innerhalb der musikalischen Dimension' (fit2_Musik) und ‚Stimmigkeit innerhalb der szenischen Dimension' (fit2_Szene) gemeinsam 83% der Varianz der ‚Stimmigkeit zwischen der musikalischen und der szenischen Dimension' (fit1) (siehe den Wert links unterhalb der entsprechenden Ellipse). Auffällig ist die Dominanz der szenischen über die musikalische Dimension (.87 vs. .28), auf welche später genauer eingegangen werden wird. Auch der Einfluss der ‚Stimmigkeit zwischen der musikalischen und der szenischen Dimension' (fit1) auf das Urteil über die künstlerische Qualität kann der Teilstruktur des Gesamtmodells in Abbildung 7 entnommen werden. 65% der Varianz des Konstruktes ‚künstlerische Qualität' können so erklärt werden. Vor der weiteren Interpretation der Ergebnisse in Hinblick auf Untersuchungsfrage 1b muss jedoch zunächst die Übereinstimmung des analytischen Modells mit den empirisch erhobenen Daten getestet werden, was sich – analog zu Pfadanalysen – mit Gütemaßen durchführen lässt. Eine Auswahl geeigneter Kriterien kann Tabelle 18 entnommen werden.

Tabelle 18: Anpassungsgüte der Teilstruktur I des Gesamtmodells über das Urteil über die künstlerische Qualität einer Opernaufführung

χ^2	df	p	χ^2/df	RMSEA	NFI	CFI
446.00	149	.000	2.99	.129	.811	.863

Anmerkungen: χ^2 = Chi-Quadrat-Wert, df = Freiheitsgrade, p = Wahrscheinlichkeit, RMSEA = Root Mean Square Error of Approximation, NFI = Normed Fit Index, CFI = Comparative Fit Index.

Bei einem Chi-Quadrat-Wert (χ^2) von 446.00 muss die Nullhypothese, nach der die empirische Kovarianzmatrix der modelltheoretischen entspricht, abgelehnt werden (p = .000). Da auch das Verhältnis zwischen χ^2 und Freiheitsgraden (df = 149) den üblicherweise auf 2.5 festgesetzten Grenzwert überschreitet (χ^2/df = 2.99), muss von einer suboptimalen Modellanpassung der Teilstruktur ausgegangen werden (Homburg & Baumgartner, 1995, S. 170).

Auch der Root Mean Square Error of Approximation (RMSEA), bei dem Werte ≤ .05 für eine hohe Anpassungsgüte sprechen, wohingegen Werte größer .10 nicht mehr akzeptiert werden können (Browne & Cudeck, 1993), lässt keineswegs auf eine hohe Übereinstimmung der Teilstruktur mit den empirischen Daten schließen (RMSEA = .13). Dies wird bestätigt durch den Normed Fit Index (NFI) und den Comparative Fit Index (CFI), der zusätzlich die Anzahl der Freiheitsgrade einbezieht. Bei beiden Kriterien deuten Werte ≥ .90 auf eine gute

Übereinstimmung zwischen Theorie und Empirie hin (Backhaus et al., 2003, S. 374f.), die in dieser Arbeit mit NFI = .81 und CFI = .86 jedoch nicht erreicht werden.

Auch das Messmodell der ‚Stimmigkeit innerhalb der szenischen Dimension' (fit2_Szene), welches dem Anhang E entnommen werden kann, kann als Ersatz für die Betrachtung der Gesamtstruktur getrennt analysiert werden.[1] Da hier ebenfalls sämtliche manifesten Variablen in hohem Ausmaß auf die latenten Konstrukte laden, können die Ergebnisse der explorativen Faktorenanalyse aus Abschnitt 4.2.3 auch für diese Teilstruktur konfirmatorisch bestätigt werden. 73% der Varianz des Konstruktes ‚fit2_Szene' können erklärt werden, wozu die latente Variable ‚Szene_Ausstattung' (.78) erheblich mehr beiträgt als ‚Chor_Szene' (.34). Wie in Tabelle 19 zu sehen ist, weisen die Gütekriterien jedoch auch in dieser Teilstruktur keineswegs auf eine hohe Anpassungsgüte hin, so dass auf weitere inhaltliche Interpretationen im Sinne der Untersuchungsfrage 1b an dieser Stelle vorerst verzichtet wird.

Tabelle 19: Anpassungsgüte der Teilstruktur II des Gesamtmodells für das Urteil über die künstlerische Qualität einer Opernaufführung

χ^2	df	p	χ^2/df	NFI	CFI	RMSEA
278.48	75	.000	3.713	.844	.879	.151

Anmerkungen: χ^2 = Chi-Quadrat-Wert, df = Freiheitsgrade, p = Wahrscheinlichkeit, NFI = Normed Fit Index, CFI = Comparative Fit Index, RMSEA = Root Mean Square Error of Approximation.

Die dritte Teilstruktur des Gesamtmodells, welche das Messmodell der ‚Stimmigkeit innerhalb der musikalischen Dimension' mit den Konstrukten ‚Orchester', ‚Solisten' und ‚Chor (Musik, fit3)' und damit eine erheblich größere Anzahl zu schätzender Parameter enthält, kann aufgrund des unzureichenden Stichprobenumfangs nicht berechnet werden.[2] Eine Darstellung des Teilmodells ohne Parameterschätzungen findet sich in Anhang F.

Obwohl eine Untersuchung der Gesamtstruktur des hierarchischen Modells im Rahmen von Strukturgleichungsmodellen aufgrund der geringen Stichprobengröße nicht möglich ist, können aus der Analyse der Teilstrukturen interessante Schlüsse gezogen werden. So können die Ergebnisse der explorativen Faktorenanalysen (vgl. Abschnitt 4.2.3) mit Hilfe der Erstellung von Messmodellen auf allen drei Ebenen konfirmatorisch bestätigt werden; lediglich zwei Items wurden entfernt. Auch können Strukturmodelle für Teilstrukturen erstellt werden, wobei inhaltliche Interpretationen der Koeffizienten im Hinblick auf die Gewichtung

[1] Bei 14 zu schätzenden Parametern dieser Teilstruktur wäre eine Stichprobe von n = 70 ausreichend, die in der vorliegenden Arbeit – wie erwähnt – erreicht wird.
[2] Notwendig wäre ein Stichprobenumfang von n = 155.

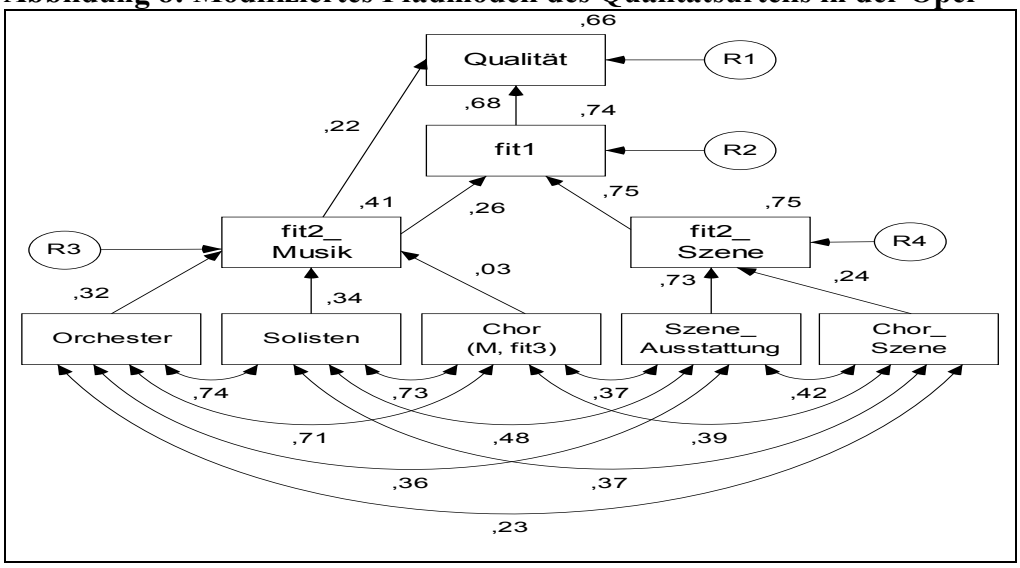
der Teilaspekte für das Publikumsurteil über die künstlerische Qualität aufgrund der mangelhaften Anpassungsgüte mit äußerster Vorsicht erfolgen müssen. Da in dieser Arbeit jedoch weniger die Teilstrukturen als vielmehr das Gesamtmodell im Mittelpunkt stehen und die Gewichtung der Teilaspekte für das Gesamturteil über die künstlerische Qualität ermittelt werden soll (Untersuchungsfrage 1b), wird im Folgenden ein erneuter Versuch der Rekonstruktion des hierarchischen Modells des Qualitätsurteils bei Opernaufführungen von Boerner (2002) mit Hilfe von Pfadanalysen erfolgen.

Pfadanalyse

Anders als Strukturgleichungsmodelle prüfen Pfadanalysen ausschließlich die Beziehungen zwischen manifesten Variablen (Bortz, 1999, S. 456). Da sie zudem keine theoretischen Annahmen über die Entstehung systematischer Messfehler treffen (Schnell et al., 2005, S. 462), reichen zu ihrer Durchführung deutlich geringere Stichproben aus, so dass mit ihrer Hilfe die Gesamtstruktur des hierarchischen Modells ermittelt werden kann. Da sich oben zeigte, dass das hierarchische Pfadmodell des Qualitätsurteils in der Oper (siehe Abbildung 5) zwar dem Generalfaktormodell (siehe Abbildung 6) deutlich überlegen ist, seine Gütekriterien jedoch auf eine suboptimale Abbildung der empirischen Daten hinweisen, liegen Modifikationen des Modells mit dem Ziel einer besseren Anpassungsgüte nahe. Es ist zu beachten, dass sich die Pfadanalyse damit von ihrem konfirmatorischen, d.h. hypothesen-testenden, Charakter entfernt und zu einem explorativen Datenanalyseinstrument wird (Backhaus et al., 2003, S. 378). Eine erneute Datenerhebung zur Überprüfung des modifizierten Modells ist demnach unbedingt notwendig. Aus Zeitgründen kann dies im Rahmen der vorliegenden Arbeit jedoch nicht erfolgen.

Nach einem Vergleich ausgewählter potentieller Modelle stellt sich das unten abgebildete modifizierte Pfadmodell als die empirischen Daten am besten abbildend dar. Wie in Abbildung 8 zu sehen ist, wurden ein direkter Einfluss der ‚Stimmigkeit innerhalb der musikalischen Dimension' (fit2_Musik) auf das Urteil über die künstlerische Qualität sowie Korrelationen zwischen den Konstrukten der dritten Ebene zugelassen.

Abbildung 8: Modifiziertes Pfadmodell des Qualitätsurteils in der Oper



Anmerkungen: Rechtecke präsentieren manifeste Variablen, Pfeile stellen ihre Zusammenhänge dar. ‚R' präsentiert eine Residualvariable. Werte am rechten oberen Rand der manifesten Variablen stellen die erklärte Gesamtvarianz dar, Werte bei den Pfeilen präsentieren die standardisierten ß-Koeffizienten.

Verschiedene Werte der Anpassungsgüte des modifizierten Pfadmodells können Tabelle 20 entnommen werden, wobei zur besseren Vergleichbarkeit die Werte des ursprünglichen hierarchischen Modells aus Abbildung 5 ebenfalls enthalten sind. Sämtliche Werte verbessern sich; die Differenz zwischen der mit Hilfe der geschätzten Parameter berechenbaren modelltheoretischen Korrelationsmatrix und der empirischen Korrelationsmatrix wird minimiert. Die Nullhypothese, nach der die empirische mit der modelltheoretischen Kovarianzmatrix übereinstimmt, muss nach dem Chi-Quadrat-Test jedoch weiterhin abgelehnt werden ($p = .000$). Da dieser wie oben erwähnt jedoch sehr sensibel auf die Anzahl der zu schätzenden Parameter reagiert, die in diesem Fall vergleichsweise hoch ist, ist der Quotient aus dem Chi-Quadrat-Wert und der Anzahl der Freiheitsgrade wesentlich aussagekräftiger. Mit $\chi^2/df = 4.36$ liegt dieser jedoch ebenfalls deutlich über dem in der Literatur genannten Grenzwert von 2.5, so dass die Anpassung auch des modifizierten Modells nicht optimal ist. Da jedoch sowohl der NFI als auch der CFI, die die Übereinstimmung zwischen Theorie und Empirie im Vergleich zu Modellen mit weniger Parametern (‚Basismodellen') messen und sich zwischen 0 und 1 bewegen, über dem Grenzwert von .90 liegen, kann insgesamt von einer akzeptablen Anpassungsgüte des modifizierten Modells ausgegangen werden (Backhaus et al., 2003, S. 372f.).

Tabelle 20: Anpassungsgüte des Modellvergleichs

	χ^2	df	p	χ^2/df	RMSEA	NFI	CFI	AIC	ECVI
modifiziertes hierarchisch. Modell (Abbildung 8)	74.17	17	.000	4.36	.168	.903	.921	148.17	1.25
hierarchisches Modell (Abbildung 5)	320.07	28	.000	11.43	.296	.582	.595	372.07	3.13

Anmerkungen: χ^2 = Chi-Quadrat-Wert, df = Freiheitsgrade, p = Wahrscheinlichkeit, RMSEA = Root Mean Square Error of Approximation, NFI = Normed Fit Index, CFI = Comparative Fit Index, AIC = Akaike Information Criterion, ECVI = Expected Cross-Validation Index.

Mit Hilfe einer Pfadanalyse kann somit die Gesamtstruktur des Publikumsurteils über die künstlerische Qualität in der Oper über die theoretisch angenommenen Ebenen rekonstruiert werden (siehe Abbildung 8), wobei sich die Gewichtung der Teilaspekte für das Gesamturteil über die künstlerische Qualität aus den standardisierten ß-Koeffizienten entnehmen lässt. 66% der Varianz der abhängigen Variable ‚künstlerische Qualität' werden erklärt, von denen ß = .68 auf die ‚Stimmigkeit zwischen der musikalischen und der szenischen Dimension' (fit1) und ß = .22 auf die ‚Stimmigkeit innerhalb der musikalischen Dimension' (fit2_Musik) entfallen. Ein direkter Einfluss der ‚Stimmigkeit innerhalb der szenischen Dimension' (fit2_Szene) auf die künstlerische Qualität besteht nicht. Auffallend bei der Betrachtung der ‚Stimmigkeit zwischen der musikalischen und der szenischen Dimension' (fit1), bei der 74% der Gesamtvarianz erklärt werden, ist die Dominanz des Einflusses der ‚Stimmigkeit innerhalb der szenischen Dimension' (fit2_Szene, ß = .75) über die ‚Stimmigkeit innerhalb der musikalischen Dimension' (fit2_Musik, ß = .26).[1] Ebenfalls bemerkenswert ist der geringe Einfluss des Konstruktes ‚musikalische Leistung und Zusammenspiel innerhalb des Chores' (Chor_M, fit3) der dritten Ebene auf die ‚Stimmigkeit innerhalb der musikalischen Dimension' (fit2_Musik), wohingegen sich der Einfluss des Orchesters (ß = .32) kaum von dem der Solisten (ß = .34) unterscheidet. Unterschiedlich stark hingegen wirken die Konstrukte ‚Szene_Ausstattung' und ‚szenische Leistung des Chores' (Chor_Szene) der dritten Ebene auf die ‚Stimmigkeit innerhalb der szenischen Dimension' (fit2_Szene), wobei der Einfluss des erstgenannten Konstruktes deutlich überwiegt (ß = .73 vs. ß = .24).

Da neben den direkten ebenfalls indirekte Effekte bestehen, erscheint zur Ermittlung der Gewichtung der Konstrukte der zweiten und dritten Ebene des analytischen Modell Boerners (2002) auf das Urteil über die künstlerische Qualität auch die Betrachtung der totalen Effekte,

[1] Dies deutete bereits die Analyse der Interkorrelationen der Skalen an (vgl. Abschnitt 4.2.2). So korreliert die Skala ‚Stimmigkeit zwischen der musikalischen und der szenischen Dimension' deutlich stärker mit der ‚Stimmigkeit innerhalb der szenischen Dimension' ($r = .81$, $p \leq 0.01$) als mit der ‚Stimmigkeit innerhalb der musikalischen Dimension' ($r = .50$, $p \leq 0.01$).

die sich aus deren Addition ergeben, lohnend. So wirkt die ‚Stimmigkeit innerhalb der musikalischen Dimension' (fit2_Musik) sowohl direkt (ß = .22) als auch indirekt über die ‚Stimmigkeit zwischen Musik und Szene' (fit1, ß = .26 x .68) auf die künstlerische Qualität, welches einen Gesamteffekt von ß = .22 + .26 x .68 = .40 ergibt. Bei einem Blick auf die totalen Effekte der verschiedenen Konstrukte auf das Urteil über die künstlerische Qualität zeigt sich, dass die ‚Stimmigkeit zwischen Musik und Szene' (fit1) mit der größten Gewichtung in das Urteil über die künstlerische Qualität eingeht (ß = .68). Die Plätze zwei und drei belegen die ‚Stimmigkeit innerhalb der szenischen Dimension' (fit2_Szene, ß = .52) sowie die ‚Stimmigkeit innerhalb der musikalischen Dimension' (fit2_Musik, ß = .40), wobei der Einfluss des Konstruktes ‚Szene_Ausstattung' nur knapp darunter liegt (ß = .38). Mit erheblich geringerem Gewicht gehen die Solisten (ß = .14), das Orchester (ß = .13) sowie die szenische Leistung des Chores (ß = .13) jeweils für sich genommen in das Urteil des Publikums ein. Eine vollständige Übersicht über die standardisierten totalen Effekte bietet Tabelle 21.

Tabelle 21: Standardisierte totale Effekte des modifizierten Pfadmodells

unabhängig / abhängig	Chor_ Szene	Szene_ Ausstat-tung	Chor (Musik, fit3)	Orches-ter	Solisten	fit2_ Szene	fit2_ Musik	fit1
fit2_Szene	.24	.73	.00	.00	.00	.00	.00	.00
fit2_Musik	.00	.00	.03	.33	.34	.00	.00	.00
fit1	.18	.55	.01	.08	.09	.75	.26	.00
Qualität	.13	.38	.01	.13	.14	.52	.40	.68

5.2 Ergebnisse zu Untersuchungsfrage 2: Unterschiede zwischen Laien und Experten in der Bewertung der künstlerischen Qualität einer Opernaufführung

Nach Beantwortung der Untersuchungsfrage 1, was Gegenstand der Bewertung künstlerischer Qualität im Musiktheater ist, werden im Folgenden die Unterschiede zwischen den Qualitätsurteilen von Laien und Experten untersucht. Zur Beantwortung der Untersuchungsfrage 2a, die Unterschiede zwischen Laien und Experten in Bezug auf die Differenziertheit und Gewichtung der Teilaspekte erfragt, werden im folgenden Abschnitt Faktoren- und Pfadanalysen getrennt nach Laien und Experten gerechnet. Anschließend wird Untersuchungsfrage 2b (unterschiedliche Beurteilung der Qualität der Teilaspekte) anhand von T-Tests zu beantworten versucht (Abschnitt 5.2.2). In Abschnitt 5.2.3 geht es um die Höhe der Übereinstimmung (Interrater-Reliabilität) der Urteile innerhalb der beiden Beurteilergruppen (Untersuchungsfrage 2c), die im Rahmen von Intraklassenkorrelationen überprüft wird.

5.2.1 Untersuchungsfrage 2a: Unterschiede zwischen Laien und Experten hinsichtlich Differenziertheit und Gewichtung

Wie in Abschnitt 4.1.4 erläutert, wurde die Unterscheidung Laie – Experte anhand von vier Items getroffen: Neben dem Item ‚Wie groß ist Ihre Erfahrung in Bezug auf die Oper?' sind dies ‚Wie oft pro Jahr besuchen Sie eine Opernaufführung (im Durchschnitt)?', ‚Beherrschen Sie ein Instrument?' und ‚Wie gut haben Sie das aufgeführte Stück bereits gekannt?'.[1]

Führt man eine getrennte Faktorenanalyse für die erste Ebene des analytischen Modells (Boerner, 2002) durch, bei der es um die Stimmigkeit zwischen der musikalischen und der szenischen Dimension geht, ergeben sich kaum Unterschiede zwischen Laien und Experten: Die Varianz, die erklärt werden kann, liegt bei den Experten etwas höher (81% vs. 76%), außerdem sind die Faktorladungen in der Expertenstichprobe geringfügig höher, d.h. der extrahierte Faktor liefert einen etwas größeren Erklärungsbeitrag für die Variablen. Ähnliches gilt für die zweite Ebene, wo es um die Stimmigkeit innerhalb der jeweiligen Dimensionen geht. Auch hier ist die erklärte Gesamtvarianz in der Gruppe der Experten geringfügig höher als in der Teilstichprobe der Laien (72% vs. 70%).[2]

Deutliche Unterschiede hingegen zeigen sich bei separaten Faktorenanalysen auf der dritten Ebene des analytischen Modells (Boerner, 2002). Bei den Laien lassen sich – ohne Vorgabe einer Anzahl von Faktoren – sieben Faktoren ziehen, die 86% der Gesamtvarianz erklären. Von diesen lassen sich sechs Faktoren inhaltlich interpretieren: ‚Leistung des Orchesters und Zusammenspiel innerhalb des Orchesters sowie Leistung des Dirigenten', ‚szenische Leistung des Chores, Ausstattung (Einzelaspekte)', ‚musikalische Leistung des Chores', ‚musikalische und szenische Leistung des Solisten', ‚Zusammenspiel innerhalb des Chores' sowie ‚Zusammenspiel der Solisten untereinander' (Aufzählung in der Reihenfolge ihrer Varianzaufklärung). Bei den Experten lassen sich sechs Faktoren extrahieren: ‚Leistung des Orchesters und Zusammenspiel innerhalb des Orchesters sowie Leistung des Dirigenten', ‚musikalische Leistung und Zusammenspiel innerhalb des Chores sowie szenische Leistung des Solisten', ‚szenische Leistung des Chores', ‚Ausstattung (Einzelaspekte)', ‚Zusammenspiel innerhalb der Solisten' und ‚musikalische Leistung des Solisten'. Allerdings gibt es – analog zur Betrachtung des gesamten Publikums – auch bei getrennter Betrachtung Items, die auf mehr als einem Faktor laden bzw. auf einem anderen als erwartet.

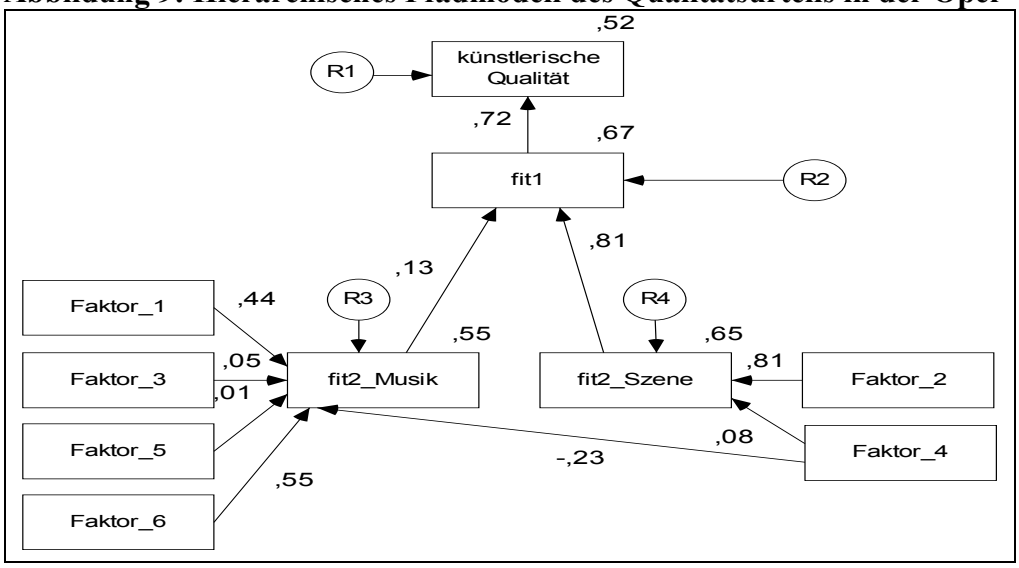
[1] Die Kennziffern der Items lauten RS4, mRS5, mRS7 und RS8; Item RS4 wurde bei der Indexbildung doppelt gewertet (vgl. Abschnitt 4.1.4).
[2] Eine Auflistung der extrahierten Faktoren sowie der Faktorladungen der einzelnen Items für alle drei Ebenen getrennt nach Experten und Laien befindet sich in Anhang M.

In beiden Gruppen lässt sich bei getrennter Betrachtung also jeweils ein Faktor mehr ziehen, als dies bei gemeinsamer Faktorenanalyse der Fall ist; dies war zumindest für die Laien nicht zu erwarten. Anders als bei der Studie von Boerner und Renz (in Vorbereitung), wo bei den Laien nur zwei Faktoren extrahiert werden konnten, sind diese in der vorliegenden Untersuchung in der Lage, nicht nur die physisch unterscheidbaren Konstrukte, sondern auch deren unterschiedliche Funktionen getrennt wahrzunehmen.[1] Fehlendes Wissen bzw. mangelnde Erfahrung, welche die Laien an der simultanen Wahrnehmung der Vielzahl von Stimuli einer Opernaufführung hindern könnten, scheiden als Erklärung für den Grad der Differenziertheit der Wahrnehmung demnach vorläufig aus. Dafür trennen die Laien nicht zwischen der szenischen Leistung des Chores und der Ausstattung (Moser, 2005, S. 51), die gemeinsam den zweiten Faktor bilden, was aufgrund der wenig lebendigen Rolle des Chores in der bewerteten Aufführung nicht überrascht. Bei den Experten hingegen wird die Ausstattung getrennt wahrgenommen, dafür differenzieren diese nicht zwischen der musikalischen Leistung des Chores, dem Zusammenspiel innerhalb des Chores und der szenischen Leistung des Solisten (Faktor 2). Da Renz (2006) zeigen konnte, dass sich sowohl die wahrgenommenen Teilaspekte eines Konstruktes je nach Werk ändern als auch die ‚Unterkonstrukte' vom Publikum werkabhängig unterschiedlich wahrgenommen werden (Renz, 2006, S. 55), kann dieses Ergebnis im Rahmen der vorliegenden Arbeit nicht näher interpretiert werden. Es bleibt festzuhalten, dass die Laien entgegen den Erwartungen zu einer ähnlich differenzierten Wahrnehmung der künstlerischen Qualität einer Aufführung im Musiktheater fähig sind wie die Experten und dass die Differenziertheit der Wahrnehmung demnach nicht allzu sehr von selbsteingeschätztem Fachwissen und Erfahrung abhängig sein kann.

Zur Überprüfung, ob Unterschiede zwischen Laien und Experten in der Gewichtung der Teilaspekte des Urteils über die künstlerische Qualität in der Oper bestehen, werden im Folgenden separate Pfadanalysen gerechnet, die auf den getrennt erfolgten Faktorenanalysen (siehe oben) aufbauen. Abbildungen 9 und 10 zeigen die hierarchischen Pfadmodelle des Qualitätsurteils in der Oper getrennt nach Laien und Experten.

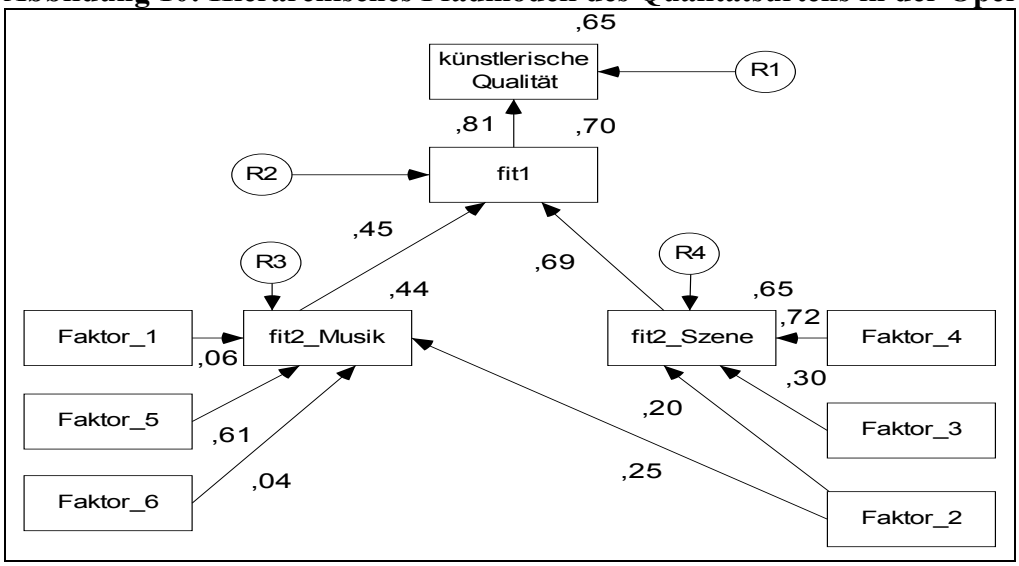
[1] Im Gegensatz zur gemeinsamen Betrachtung von Laien und Experten gilt dies nicht nur für den Chor, sondern auch für die Solisten.

Abbildung 9: Hierarchisches Pfadmodell des Qualitätsurteils in der Oper – Laien



Anmerkungen: Rechtecke präsentieren manifeste Variablen, Pfeile stellen ihre Zusammenhänge dar. ‚R' präsentiert eine Residualvariable. Werte am rechten oberen Rand der manifesten Variablen stellen die erklärte Gesamtvarianz dar, Werte bei den Pfeilen präsentieren die standardisierten ß-Koeffizienten. Faktor 1 = Leistung des Orchesters und Zusammenspiel innerhalb des Orchesters sowie Leistung des Dirigenten, Faktor 2 = szenische Leistung des Chores, Ausstattung (Einzelaspekte), Faktor 3 = musikalische Leistung des Chores, Faktor 4 = musikalische und szenische Leistung des Solisten, Faktor 5 = Zusammenspiel innerhalb des Chores, Faktor 6 = Zusammenspiel der Solisten untereinander.

Abbildung 10: Hierarchisches Pfadmodell des Qualitätsurteils in der Oper – Experten



Anmerkungen: Rechtecke präsentieren manifeste Variablen, Pfeile stellen ihre Zusammenhänge dar. ‚R' präsentiert eine Residualvariable. Werte am rechten oberen Rand der manifesten Variablen stellen die erklärte Gesamtvarianz dar, Werte bei den Pfeilen präsentieren die standardisierten ß-Koeffizienten. Faktor 1 = Leistung des Orchesters und Zusammenspiel innerhalb des Orchesters sowie Leistung des Dirigenten, Faktor 2 = musikalische Leistung und Zusammenspiel innerhalb des Chores sowie szenische Leistung des Solisten, Faktor 3 = szenische Leistung des Chores, Faktor 4 = Ausstattung (Einzelaspekte), Faktor 5 = Zusammenspiel innerhalb der Solisten, Faktor 6 = musikalische Leistung des Solisten.

Zur Beantwortung des zweiten Teils der Untersuchungsfrage 2a, welche Unterschiede zwischen Laien und Experten in Bezug auf die Gewichtung der Teilsaspekte bestehen, ist die Betrachtung der standardisierten totalen Effekte, die sich aus der Addition der direkten und

indirekten Effekte ergeben, geeignet. Da die Faktorenanalyse in beiden Beurteilergruppen jeweils unterschiedliche Faktoren ergab (siehe oben), sind die Ergebnisse getrennt dargestellt (siehe Tabellen 22 bzw. 23).

Tabelle 22: Standardisierte totale Effekte des hierarchischen Pfadmodells - Laien

unabhängig / abhängig	Faktor 1	Faktor 2	Faktor 3	Faktor 4	Faktor 5	Faktor 6	fit2_ Szene	fit2_ Musik	fit1
fit2_Szene	.00	.81	.00	.08	.00	.00	.00	.00	.00
fit2_Musik	.44	.00	.05	-.23	.01	.55	.00	.00	.00
fit1	.06	.65	.01	.03	.00	.07	.81	.13	.00
Qualität	.04	.47	.01	.02	.00	.05	.58	.09	.72

Anmerkungen: Faktor 1 = Leistung des Orchesters und Zusammenspiel innerhalb des Orchesters sowie Leistung des Dirigenten, Faktor 2 = szenische Leistung des Chores, Ausstattung (Einzelaspekte), Faktor 3 = musikalische Leistung des Chores, Faktor 4 = musikalische und szenische Leistung des Solisten, Faktor 5 = Zusammenspiel innerhalb des Chores, Faktor 6 = Zusammenspiel der Solisten untereinander.

Tabelle 23: Standardisierte totale Effekte des hierarchischen Pfadmodells - Experten

unabhängig / abhängig	Faktor 1	Faktor 2	Faktor 3	Faktor 4	Faktor 5	Faktor 6	fit2_ Szene	fit2_ Musik	fit1
fit2_Szene	.00	.20	.30	.72	.00	.00	.00	.00	.00
fit2_Musik	.06	.25	.00	.00	.61	.04	.00	.00	.00
fit1	.03	.25	.21	.50	.27	.02	.69	.45	.00
Qualität	.02	.20	.17	.40	.22	.01	.56	.36	.81

Anmerkungen: Faktor 1 = Leistung des Orchesters und Zusammenspiel innerhalb des Orchesters sowie Leistung des Dirigenten, Faktor 2 = musikalische Leistung und Zusammenspiel innerhalb des Chores sowie szenische Leistung des Solisten, Faktor 3 = szenische Leistung des Chores, Faktor 4 = Ausstattung (Einzelaspekte), Faktor 5 = Zusammenspiel innerhalb der Solisten, Faktor 6 = musikalische Leistung des Solisten.

Bei einem Vergleich der totalen Effekte der verschiedenen Konstrukte auf das Publikumsurteil über die künstlerische Qualität fällt auf, dass die ‚Stimmigkeit zwischen der musikalischen und der szenischen Dimension' (fit1) jeweils mit der größten Gewichtung in das Urteil über die künstlerische Qualität eingeht (Laien: ß = .72, Experten ß = .81). Den zweiten Platz belegt in beiden Beurteilergruppen die ‚Stimmigkeit innerhalb der szenischen Dimension' (fit2_Szene) (ß = .58 bzw. ß = .56). Ebenfalls nur leichte Unterschiede bestehen bezüglich des Konstruktes, welches an dritter Stelle in das Urteil über die künstlerische Qualität eingeht: Bei den Laien ist dies der Faktor, der sich aus der szenischen Leistung des Chores und der Ausstattung (Einzelaspekte) zusammensetzt (ß = .47), wohingegen die Experten den Faktor ‚Ausstattung (Einzelaspekte)' separat sehen und diesem ein etwas geringeres Gewicht einräumen (ß = .40). Deutliche Unterschiede zwischen Laien und Experten bestehen in Hinblick auf das Konstrukt ‚Stimmigkeit innerhalb der musikalischen Dimension' (fit2_Musik), die bei den Experten an vierter Stelle in das Urteil über die künstlerische Qualität eingeht (ß = .36), wohingegen sie in der Teilstichprobe der Laien kaum

eine Rolle für das Urteil spielt (ß = .09). Die Vermutung, dass Fachwissen und Erfahrung speziell zur Beurteilung der Stimmigkeit innerhalb der musikalischen Dimension nötig seien, die die Laiengruppe nach eigener Einschätzung nicht bzw. nur in geringem Ausmaß haben, und dass das Konstrukt bei den Laien daher weniger wichtig für das Gesamturteil über die künstlerische Qualität sei, scheint bestätigt. Ebenfalls deutlich wird dies bei Betrachtung des Einflusses der Konstrukte der zweiten Ebene der Pfadmodelle auf die ‚Stimmigkeit zwischen der musikalischen und der szenischen Dimension' (fit1, Ebene 1). Überwiegt in beiden Beurteilergruppen der Einfluss der ‚Stimmigkeit innerhalb der szenischen Dimension' (fit2_Szene, ß = .81 bzw. ß = .69), ist das Verhältnis zwischen der ‚Stimmigkeit innerhalb der szenischen Dimension' (fit2_Szene) und der ‚Stimmigkeit innerhalb der musikalischen Dimension' (fit2_Musik) in der Expertengruppe erheblich ausgeglichener. Insgesamt können 52% (Laien) bzw. 65% (Experten) der Gesamtvarianz erklärt werden.

Zusammenfassend lässt sich in Hinblick auf Untersuchungsfrage 2a – welche Unterschiede zwischen der Teilstichprobe der Laien und der Experten in Bezug auf die Differenziertheit und Gewichtung der Teilaspekte des Urteils über die künstlerische Qualität einer Opernaufführung bestehen – konstatieren, dass sich sowohl Differenziertheit als auch Gewichtung der Teilaspekte in beiden Gruppen insgesamt nur wenig unterscheiden, so dass beides nicht allzu sehr von selbsteingeschätztem Fachwissen und Erfahrung abhängig sein kann.

5.2.2 Untersuchungsfrage 2b: Unterschiede zwischen Laien und Experten in der Bewertung der Teilaspekte

Zur Klärung der Untersuchungsfrage 2b, ob die künstlerische Qualität der Teilaspekte einer Opernaufführung von Laien und Experten unterschiedlich beurteilt wird, werden im Folgenden T-Tests durchgeführt. Diese dienen der Überprüfung, ob eventuell beobachtete Unterschiede der Mittelwerte zwischen Laien und Experten zufällig entstanden sind oder eher systematische Ursachen haben, also auch in der Grundgesamtheit voneinander verschieden sind (Brosius, 2006, S. 471f.).[1] Eine Auflistung der T-Werte sämtlicher Skalen, die anhand der theoretischen Konstrukte (Boerner, 2002) gebildet wurden, findet sich in Tabelle 24.[2]

[1] Auf sehr ähnliche Weise funktionieren auch die Varianzanalysen, die zur Ermittlung von eventuellen Unterschieden der Bewertung künstlerischer Qualität je nach Zeitpunkt des Ausfüllens des Fragebogens durchgeführt wurden (vgl. Abschnitt 4.1.2). Im Unterschied zu T-Tests lassen sich mit diesen jedoch mehr als zwei Mittelwerte gleichzeitig testen.

[2] Eine Erläuterung zur Entstehung der Skalen, die Reliabilitäten sowie eine Auflistung der Items, welche die Skalen jeweils bilden, ist in Abschnitt 4.2.1 zu finden.

Tabelle 24: Vergleich der Qualitätsunterschiede von Experten und Laien

Skala	T-Wert	Mittelwert		Standardabweichung		Stichprobengröße	
		Laien	Experten	Laien	Experten	Laien	Experten
kQ	1.400	4.00	3.74	.886	1.089	57	56
fit1	-0.577	3.98	4.08	.883	.850	50	50
fit2_M	1.285	4.39	4.22	.673	.715	53	56
fit2_S	-0.683	3.79	3.92	.960	.967	52	56
MO	3.365**	4.71	4.23	.460	.842	37	49
MSo	0.789	4.46	4.36	.502	.702	49	54
MChor	2.213*	4.58	4.27	.461	.905	45	57
SSo	0.744	4.16	4.03	.648	1.083	54	56
SChor	-1.208	3.51	3.77	1.082	1.137	50	56
SA	-0.356	3.70	3.77	1.138	.980	56	56
fit3_Chor	1.658	4.50	4.30	.436	.753	45	54
fit3_So	1.806	4.29	4.04	.647	.792	53	57

Anmerkungen: *p ≤ .05, **p ≤ .001, *** p ≤ .000. kQ = künstlerische Qualität, fit1 = Stimmigkeit zwischen der musikalischen und der szenischen Dimension, fit2_M = Stimmigkeit innerhalb der musikalischen Dimension, fit2_S = Stimmigkeit innerhalb der szenischen Dimension, MO = musikalische Leistung und Zusammenspiel innerhalb des Orchesters, MD = Leistung des Dirigenten, MSo = musikalische Leistung des Solisten, MChor = musikalische Leistung des Chores, SSo = szenische Leistung des Solisten, SChor = szenische Leistung des Chores, SA = Szene_Ausstattung (Einzelaspekte), fit3_Chor = Zusammenspiel innerhalb des Chores, fit3_So = Zusammenspiel der Solisten untereinander.

Die Vermutung, nach der Experten aufgrund ihres Fachwissens und der größeren Erfahrung, die ihnen mehr Vergleiche zu anderen Produktionen und Opernhäusern ermöglichen sowie ihre Erwartungen steigern könnten, kritischer über die künstlerische Qualität einer Opernaufführung urteilten (Boerner & Renz, in Vorbereitung), wird nicht bestätigt. Lediglich bei den Skalen ‚musikalische Leistung des Orchesters' und ‚musikalische Leistung des Chores' bestehen signifikante Mittelwertsdifferenzen, d.h. die Experten bewerten die künstlerische Qualität schlechter als die Laien.[1] Die in Abschnitt 5.1.2 erwähnte Vermutung, speziell zur Beurteilung der musikalischen Dimension seien Fachwissen und Erfahrung notwendig, da sich für Laien vermutlich jedes professionelle Orchester und jeder Chor ‚gut' anhörten, kann somit vorerst bestätigt werden. Im Unterschied dazu sind bei den restlichen Skalen eventuell bestehende Unterschiede der Mittelwerte nicht signifikant. Ein vergleichbares Resultat ergibt die Untersuchung von Boerner und Renz (in Vorbereitung). Sie begründen die nur geringen und nicht signifikanten Mittelwertsunterschiede zwischen Laien und Experten damit, dass nahezu alle Opernbesucher zumindest etwas Erfahrung mitbringen und damit die notwendige Vertrautheit besitzen, die sie zur Bewertung der künstlerischen Qualität einer Opernaufführung für notwendig erachten. Ähnliche Ergebnisse, d.h. keine

[1] Gemäß der verwendeten Item-Kodierung ist die Bewertung der künstlerischen Qualität dabei umso positiver, je höher der Mittelwert ist.

signifikanten Unterschiede zwischen Laien und Experten, finden sich ebenfalls in der Musikpsychologie (u.a. Winter, 1993; Byo & Brooks, 1994; Thompson, 2006).

Auffallend bei weiterer Betrachtung der Tabelle 24 sind die Standardabweichungen der Skalenmittelwerte, die bei den Experten deutlich größer sind als bei den Laien.[1] Die Übereinstimmung der Urteile innerhalb der verschiedenen Beurteilergruppen (Untersuchungsfrage 2c) wird daher im nächsten Abschnitt explizit behandelt.

5.2.3 Untersuchungsfrage 2c: Übereinstimmung der Urteile jeweils innerhalb der Beurteilergruppen

Zur Beantwortung der Untersuchungsfrage 2c wird im Folgenden getestet, wie hoch ist die Übereinstimmung (Interrater-Reliabilität) der Urteile innerhalb der Laien- und der Expertengruppe ist. Es soll also an dieser Stelle überprüft werden, ob die Einschätzung der Konstrukte durch jeden einzelnen Befragten zuverlässig oder durch eine hohe Subjektivität gekennzeichnet ist. Da Unterschiede zwischen Laien und Experten vermutet werden, wird deren Übereinstimmung bzgl. der Beurteilung im Folgenden separat untersucht.

Um das am besten geeignete Zuverlässigkeitsmaß auszuwählen, muss an erster Stelle das Skalenniveau der Daten betrachtet werden. Da im vorliegenden Fall eine Likert-Skala zur Anwendung kam (vgl. Abschnitt 3.1.1) und die einzelnen Skalen-Punkte diskrete Ausprägungen einer kontinuierlichen latenten Variablen darstellen, wurde das Skalenniveau der Intervallskala erreicht (Schnell et al., 2005, S. 143). Aus diesem Grund ist es – im Unterschied zu nominalskalierten Kategoriensystemen – für eine hohe Zuverlässigkeit der Urteile nicht notwendig, dass alle Befragten (‚Rater') genau gleich urteilen, d.h. in der Beurteilung der künstlerischen Qualität der jeweiligen Konstrukte im Allgemeinen exakt übereinstimmen. Von einer hohen Zuverlässigkeit kann bei intervallskalierten Ratingdaten auch dann ausgegangen werden, wenn ein Konstrukt von den verschiedenen Ratern *ähnliche* Werte erhält (Wirtz & Caspar, 2002, S. 35).[2] Eine demnach geeignete Methode zur Reliabilitätsbestimmung stellt die Intraklassenkorrelation (ICC) dar, die als Korrelation eines beliebigen Raterpaares definiert wird (Wirtz & Caspar, 2002, S. 232) und demnach Informationen über die Stärke des Zusammenhangs zwischen Ratern liefert. Die ICC kann

[1] Ausnahmen bilden die Skalen ‚Stimmigkeit zwischen der musikalischen und der szenischen Dimension' (fit1) und ‚Szene_Ausstattung' (Einzelaspekte), wo die Skalenmittelwerte der Laien eine größere Standardabweichung aufweisen.

[2] So ist von einer hohen Übereinstimmung auch dann auszugehen, wenn der Großteil der Befragten ein Item mit „trifft voll und ganz zu" oder mit „trifft eher zu" beantwortet und andere Antwortmöglichkeiten nur von wenigen gewählt werden.

Werte zwischen 0 und 1 annehmen, wobei ein Wert von 0 bedeutet, dass kein Zusammenhang zwischen den Urteilen der Rater besteht, und ein Wert von 1 einen perfekten Zusammenhang, d.h. eine exakte Übereinstimmung der Urteile, indiziert. Die ICC wird somit ähnlich interpretiert wie die Produkt-Moment-Korrelation, welche zu den am häufigsten verwendeten Korrelationsmaßen zählt (vgl. Abschnitt 4.2.2). Anders als bei dieser kann bei Verwendung der ICC jedoch frei entschieden werden, ob Mittelwertsunterschiede zwischen den Befragten in die Berechnung der Güte des Zusammenhangs eingehen (sog. unjustierte Maße) oder ob diese in der Korrelationsschätzung nicht abgebildet werden sollen (justierte Maße).[1] [2] Gemäß dem unjustierten Modell kann die ICC den Wert 1 also nur erreichen, wenn die Varianzen für die Rater gleich sind und zusätzlich die Gleichheit der Mittelwerte gegeben ist. ICC_{unjust} stellt demnach das strengere Modell dar und wird im Folgenden verwendet werden.[3] Tabelle 25 zeigt die Intraklassenkorrelationen der theoretisch angenommenen Konstrukte (Boerner, 2002).[4]

Tabelle 25: Berechnung der Intraklassenkorrelationen

Skala	ICC_{unjust}		Stichprobengröße	
	Laien	Experten	Laien	Experten
kQ	.873***	.893***	57	56
fit1	.902***	.933***	50	50
fit2_M	.814***	.827***	53	56
fit2_S	.940***	.952***	52	56
MO	.960***	.968***	37	49
MSo	.844***	.917***	49	54
MChor	.926***	.965***	45	57
SSo	.554***	.918***	54	56
SChor	.939***	.940***	50	56
SA	.899***	.872***	56	56
fit3_Chor	.714***	.826***	45	54
fit3_So	.720***	.798***	53	57

Anmerkungen: *** $p \leq .000$. kQ = künstlerische Qualität, fit1 = Stimmigkeit zwischen der musikalischen und der szenischen Dimension, fit2_M = Stimmigkeit innerhalb der musikalischen Dimension, fit2_S = Stimmigkeit innerhalb der szenischen Dimension, MO = musikalische Leistung und Zusammenspiel innerhalb des Orchesters, MSo = musikalische Leistung des Solisten, MChor = musikalische Leistung des Chores, SSo = szenische Leistung des Solisten, SChor = szenische Leistung des Chores, SA = Szene_Ausstattung (Einzelaspekte), fit3_Chor = Zusammenspiel innerhalb des Chores, fit3_So = Zusammenspiel der Solisten untereinander. Da jeweils nur die Befragten berücksichtigt wurden, die sich zu einer Beurteilung in der Lage sahen, d.h. nicht die Antwortmöglichkeiten „habe ich nicht beachtet“ bzw. „kann ich nicht beurteilen“ wählten, variieren die Stichprobengrößen je nach Konstrukt geringfügig.

[1] Die Produkt-Moment-Korrelation ist von ihrer Berechnung her immer ein justiertes Zusammenhangsmaß, d.h. Mittelwertsunterschiede gehen nicht in die Berechnung des Zusammenhangs ein (Wirtz & Caspar, 2002, S. 232).

[2] vgl. zu weiteren Vorteilen der ICC gegenüber der Produkt-Moment-Korrelation Wirtz und Caspar, 2002, S. 157f.

[3] Insbesondere, weil die Ratingwerte unabhängig vom individuellen Standard der verschiedenen Rater weiter verwendet werden sollen, erscheint ICC_{unjust} hier als geeignetes Maß. So ist von einzelnen Befragten nicht bekannt, ob sie insgesamt eher kritisch oder wohlwollend urteilen.

[4] Eine Erläuterung zur Entstehung der Skalen, Cronbachs Alpha sowie eine Auflistung der Items, welche die Skalen jeweils bilden, ist in Abschnitt 4.2.1 zu finden.

Vor der inhaltlichen Interpretation der Ergebnisse soll auf zwei Dinge speziell hingewiesen werden: Nach einem in der Literatur häufig verwendeten Kriterium werden Intraklassenkorrelationen ab 0.7 als Indiz für eine gute Reliabilität angesehen, wobei zu beachten ist, dass die ICC_{unjust} allgemein niedrigere Werte annimmt als die ICC_{just}.[1] Des Weiteren ist zu beachten, dass bei der Berechnung des ICC in der Ausgabe des Statistikprogramms jeweils zwei Maße angegeben werden: Neben der Reliabilität der Einzelratings ist dies ebenfalls die der Mittelwerte über mehrere Rater, wobei letztere – naturgemäß – stets höher ist (Wirtz & Caspar, 2002, S. 162). Da zur Beantwortung von Untersuchungsfrage 2c neben der Zuverlässigkeit der Urteile insbesondere der Vergleich der Mittelwerte zwischen der Laien- und der Expertengruppe interessiert, wird im Folgenden die Reliabilität der Mittelwerte betrachtet.

Beim Blick auf Tabelle 25 fällt auf, dass – im Unterschied zu den Studien von Renz (2006, S. 81) sowie Boerner und Renz (in Vorbereitung) – alle Koeffizienten sowohl bei den Laien als auch bei den Experten Werte deutlich über 0.7 aufweisen ($p \leq .000$); bei allen Skalen ist der Grad der Übereinstimmung der Urteile innerhalb der jeweiligen Gruppen demnach außerordentlich hoch.[2] Anders als bei Braun (1972), Boyle und Radcoy (1987) sowie Thompson und Williamon (2003) scheint die Bewertung künstlerischer Qualität in der Oper somit weit weniger unterschiedlich als zunächst vermutet. Verschiedene Gründe werden an dieser Stelle für das Ergebnis verantwortlich gemacht: So könnten bei den Laien fehlendes Fachwissen bzw. mangelnde Erfahrung eine eher unkritische Haltung hervorbringen, die in einheitlichen Bewertungskriterien bzw. einer übereinstimmend hohen Bewertung der künstlerischen Qualität resultiert.[3] Als ein weiterer Grund für die hohe Übereinstimmung innerhalb der Laien wird deren Unsicherheit bzgl. ihres Urteils gesehen und vermutet, dass „vor allem in nicht eindeutigen Wahrnehmungssituationen zuvor abgegebene Urteile anderer das eigene Urteil beeinflussen können“ (Behne, 1997, S. 1006, vgl. ebenfalls Furman & Duke, 1988). So könnten die (zum Großteil) sehr positiven Kritiken, die nach der Premiere der Dessauer Inszenierung der Oper „Die Zauberflöte“ in den Medien veröffentlicht wurden, zu der übereinstimmend positiven Beurteilung beigetragen haben (Eversmann, 2004, S. 168). Auch die gute Reputation sowohl des Regisseurs Johannes Felsenstein als auch des

[1] Die Werte gehen dabei umso deutlicher auseinander, je größer die Unterschiede der Mittelwerte der Befragten ausfallen, da – wie oben erläutert – nur bei der ICC_{unjust} Mittelwertsunterschiede zwischen den Befragten in die Berechnung der Güte des Zusammenhangs, d.h. in die Reliabilität, eingehen.

[2] Einzige Ausnahme bildet die Skala ‚szenische Leistung des Solisten’ (SSo) bei den Laien mit einer Intraklassenkorrelation von $ICC_{unjust} = .554$.

[3] Dies wird bestätigt durch die hohen Mittelwerte der gebildeten Skalen, die zum überwiegenden Teil mindestens 4.00 auf der fünfstufigen Likert-Skala betragen (vgl. Abschnitt 5.2.2).

Dirigenten Golo Berg könnten einen Einfluss auf die Bewertung der künstlerischen Qualität durch die Mehrzahl der Laien gehabt haben (Kreitler & Kreitler, 1980, S. 248f.; Wilson, 1985, S. 67f.; Behne, 1987). Zusätzlich könnte der nicht enden wollende Applaus nach der Aufführung in Dessau die (unsicheren) Laien in dem Sinne beeinflusst haben, „that feelings are intensified and reinforced by the reaction of their neighbours" (Eversmann, 2004, S. 171; vgl. ebenfalls Wilson, 1985, S. 65f.).[1] Demnach scheint die Bewertung künstlerischer Qualität in der Oper bei den Laien weniger eine Frage des persönlichen Geschmacks und „an individual matter" (Eversmann, 2004, S. 139), also subjektiv, sondern einheitlichen Bewertungskriterien geschuldet zu sein (vgl. Wapnick & Ekholm, 1997).[2] Wie ist mit diesem Hintergrund jedoch die Konsistenz der Urteile der Experten zu erklären, die die der Laien noch übersteigt?[3]

Anders als bei Boerner (2002), deren Studie zur Übereinstimmung von Expertenurteilen bei der Bewertung von Opernaufführungen und -häusern keine Konsistenz der Urteile ergab (2002, S. 122f.), weisen neben den Laien- auch die Expertenurteile in der vorliegenden Untersuchung große Übereinstimmungen auf (siehe Tabelle 25). Da es in der Literatur keinen allgemein akzeptierten „Kriterienkatalog" gibt, an dem sich Experten bei ihren Urteilen orientieren könnten (Moser, 2005, S. 11), und es, wie Lesle für den Bereich der Musik zeigen konnte, kein objektives Urteil geben kann (1981, S. 66), müssen andere Gründe für die erstaunliche Konsistenz der Expertenurteile gefunden werden. Als Ursache werden professionelle Qualitätsstandards vermutet, die Experten aufgrund ihres Fachwissens und ihrer Erfahrung entwickeln und anwenden, d.h. „some generally held expectations when it comes to what is considered ‚good' theatre" (Eversmann, 2004, S. 167; vgl. ebenfalls Koebner, 1993; Boorsma & van Maanen, 2003; Neuhoff & de la Motte-Haber, 2007).[4]

[1] Auch die Kodierung der Antwortmöglichkeiten „habe ich nicht beachtet" bzw. „kann ich nicht beurteilen" als ‚fehlende Werte' kommt als (Teil)Grund für die hohe Konsistenz der Urteile in Betracht, da durch diese Vorgehensweise ein Teil der abweichenden Meinungen aus der Berechnung der Intraklassenkorrelationen ausgeschlossen wurde.

[2] Erstaunlicherweise bietet die Literatur wie eingangs erwähnt keine Hinweise darauf, *welche* Bewertungskriterien dabei bei Laien zum Einsatz kommen.

[3] Einzige Ausnahme bildet die Skala ‚Szene_Ausstattung' (Einzelaspekte), wo die Konsistenz der Laienurteile die der Experten übersteigt.

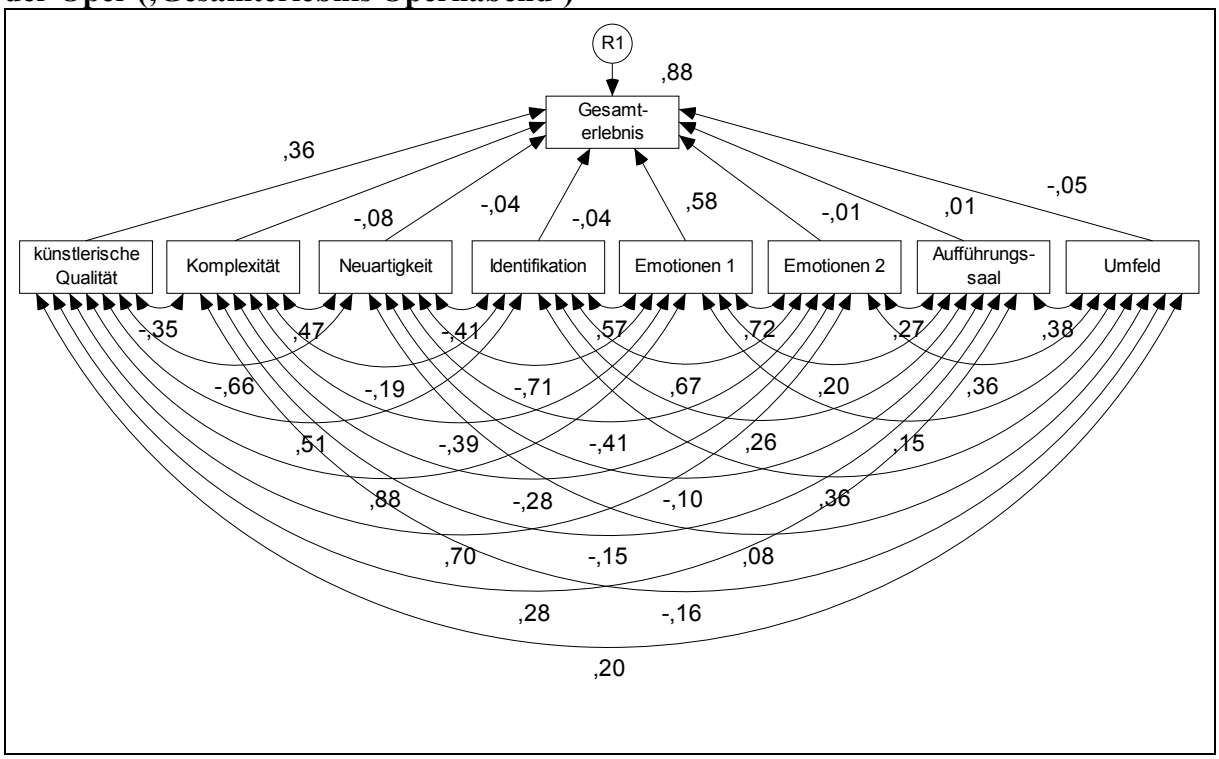
[4] Dazu ist erneut zu erwähnen, dass es sich in der zitierten Literatur bei ‚Experten' zumeist um Künstler oder Kritiker handelt, wohingegen in der vorliegenden Arbeit auch opernerfahrene Nicht-Professionelle mit Kenntnissen des aufgeführten Stückes als ‚Experten' bezeichnet werden (vgl. Abschnitt 4.1.4). Eine exakte Übertragung der Ergebnisse der Literatur auf den Kontext der vorliegenden Untersuchung ist streng genommen also nicht möglich.

5.3 Ergebnisse zu Untersuchungsfrage 3: Determinanten des Gesamturteils des Publikums über einen Abend in der Oper (‚Gesamterlebnis Opernabend')

In den folgenden Abschnitten werden nun die Determinanten des Gesamturteils des Publikums über einen Abend in der Oper (‚Gesamterlebnis Opernabend') im Mittelpunkt stehen (Untersuchungsfrage 3). So wurde bei der Vorstellung sowohl bei der Literatur zur Musikrezeption als auch zum Sprechtheater (Abschnitt 2.2.3.1 bzw. Abschnitt 2.2.3.2) deutlich, dass neben der Bewertung der künstlerischen Qualität weitere Aspekte in das Gesamturteil des Publikums über einen Abend im Konzert bzw. im Sprechtheater einfließen. Da in dieser Arbeit ähnliche Effekte auch im Musiktheater vermutet werden, wurden bei der Konstruktion des Fragebogens der vorliegenden Arbeit neben Aspekten der künstlerischen Qualität auch zusätzliche Aspekte wie bspw. die emotionale Wirkung der Aufführung auf die Zuschauer sowie Fragen zu situativen Rahmenbedingungen aufgenommen. Erläuterungen zu den Hintergründen und dem Vorgehen finden sich in den Abschnitten 2.3 und 3.1.1.

Da sich bei der Auswertung der Korrelationsmatrix zeigte (vgl. Abschnitt 4.2.2), dass die gebildeten Skalen teilweise in erheblichem Ausmaß miteinander korrelieren, erscheint eine lineare Regressionsanalyse zur Bestimmung der Determinanten des Gesamturteils des Publikums über einen Abend in der Oper wenig geeignet (Backhaus et al., 2003, S. 88f.). Um zu untersuchen, welchen Einfluss die Konstrukte ‚künstlerische Qualität', ‚Komplexität', ‚Neuartigkeit', ‚Identifikation', ‚Genussfaktor/ Ich-Beteiligung', ‚Empathie/ persönliche Assoziationen', ‚Aufführungssaal' und ‚Umfeld' auf das Publikumsurteil über das ‚Gesamterlebnis Opernabend' haben, wird daher im Folgenden eine Pfadanalyse durchgeführt (siehe Abbildung 11).

Abbildung 11: Determinanten des Gesamturteils des Publikums über einen Abend in der Oper (‚Gesamterlebnis Opernabend')



Anmerkungen: Rechtecke präsentieren manifeste Variablen, Pfeile stellen ihre Zusammenhänge dar. ‚R' präsentiert eine Residualvariable. Werte am rechten oberen Rand der manifesten Variablen stellen die erklärte Gesamtvarianz dar, Werte bei den Pfeilen präsentieren die standardisierten ß-Koeffizienten. Emotionen 1 = aus den Skalen ‚Genussfaktor' und ‚Ich-Beteiligung' im Rahmen einer explorativen Faktorenanalyse gebildeter Faktor, Emotionen 2 = aus den Skalen ‚Empathie' und ‚persönliche Assoziationen' gebildeter Faktor (vgl. Abschnitt 4.2.3).

Wie Abbildung 11 zu entnehmen ist, können 88% der Gesamtvarianz des Publikumsurteils über einen Abend in der Oper (‚Gesamterlebnis Opernabend') mit Hilfe der erarbeiteten Determinanten erklärt werden. Dabei hat das Konstrukt ‚Genussfaktor/ Ich-Beteiligung' (Emotionen 1) das größte Gewicht auf das Urteil über das Gesamterlebnis (ß = .58). An zweiter Stelle steht die künstlerische Qualität der Opernaufführung, die mit einem Gewicht von ß = .36 in das Gesamturteil eingeht. Sämtliche anderen Konstrukte haben keinen bzw. einen nur sehr geringen Einfluss. Das große Gewicht der emotionalen Wirkung der Aufführung auf das Publikumsurteil über einen Abend in der Oper überrascht insofern, als dass die Literatur – wie erwähnt – bislang „has generally focused on the actual performance itself, with the […] assumption that the listener's perception of the quality of the performance is the primary issue at hand" (Thompson, 2007, S. 20f.) und die emotionale Wirkung auf den Zuschauer wenig im Fokus stand. In Einklang hingegen steht das Resultat mit den Ergebnissen von Behr (1983), in dessen Untersuchung der Wirkungen einer Opernaufführung auf den Zuschauer ebenfalls die emotionalisierende Wirkung die größte Zustimmung vom

Publikum erhält (Behr, 1983, S. 164; vgl. ebenfalls Thompson, 2006, S. 215 für Konzertaufführungen).

5.4 Ergebnisse zu Untersuchungsfrage 4: Unterschiede zwischen Laien und Experten in der Bewertung des ‚Gesamterlebnisses Opernabend'

Nachdem mit Hilfe einer Pfadanalyse ermittelt wurde, aus welchen Teilaspekten sich das Publikumsurteil über einen Abend in der Oper zusammensetzt und mit welchem Gewicht diese jeweils eingehen (Untersuchungsfrage 3), werden nun die Bewertungen von Laien und Experten getrennt analysiert. Untersuchungsfrage 4a versucht dabei zu klären, ob Teilaspekte des Gesamturteils unterschiedlich bewertet werden. Anschließend wird die Höhe der Übereinstimmung (Interrater-Reliabilität) der Urteile innerhalb der beiden Beurteilergruppen überprüft (Untersuchungsfrage 4b). Das Kapitel schließt mit Tests zu den Unterschieden in Bezug auf die Gewichtung der Teilaspekte (Untersuchungsfrage 4c).

5.4.1 Untersuchungsfrage 4a: Unterschiede zwischen Laien und Experten in der Bewertung der Teilaspekte

Zur Klärung der Untersuchungsfrage 4a, ob die Teilaspekte des Publikumsurteils über das Gesamterlebnis eines Opernabends von Laien und Experten unterschiedlich beurteilt werden, dienen T-Tests, deren Funktionsweise Abschnitt 5.2.2 entnommen werden kann (siehe Tabelle 26).[1]

[1] zur Bildung des Indexes ‚Expertentum' vgl. Abschnitt 4.1.4

Tabelle 26: Vergleich der Bewertung des Gesamterlebnisses von Laien und Experten

Skala	T-Wert	Mittelwert		Standardabweichung		Stichprobengröße	
		Laien	Experten	Laien	Experten	Laien	Experten
Gesamturteil	.704	4.25	4.11	.868	1.120	55	56
künstlerische Qualität	1.400	4.00	3.74	.886	1.089	57	56
Komplexität	.607	1.82	1.74	.661	.747	57	56
Neuartigkeit	1.314	2.29	2.03	1.121	.983	56	56
Identifikation	.250	3.48	3.43	1.214	1.094	54	54
Emotionen 1	.518	4.07	3.99	.765	1.026	57	57
Emotionen 2	.530	3.32	3.23	.825	1.018	57	56
Aufführungs-saal	.482	4.21	4.16	.578	.608	56	55
Umfeld	.040	4.43	4.43	.596	.548	54	53

Anmerkungen: Emotionen 1 = aus den Skalen ‚Genussfaktor' und ‚Ich-Beteiligung' im Rahmen einer explorativen Faktorenanalyse gebildeter Faktor, Emotionen 2 = aus den Skalen ‚Empathie' und ‚persönliche Assoziationen' gebildeter Faktor (vgl. Abschnitt 4.2.3). Da jeweils nur die Befragten berücksichtigt wurden, die sich zu einer Beurteilung in der Lage sahen, d.h. nicht die Antwortmöglichkeiten „habe ich nicht beachtet" bzw. „kann ich nicht beurteilen" wählten, variieren die Stichprobengrößen je nach Konstrukt geringfügig.

Wie anhand Tabelle 26 ersichtlich wird, bewerten die Experten die Teilaspekte durchgängig negativer als die Laien. Einzige Ausnahme bildet die Skala ‚Umfeld', bei der sich die Mittelwerte nicht unterscheiden. Da sämtliche T-Werte das vorgegebene Signifikanzniveau von 0.05 jedoch deutlich überschreiten, muss von einem zufälligen Zustandekommen der Mittelwertsunterschiede ausgegangen werden; die Nullhypothese kann nicht zurückgewiesen werden. Dies überrascht insbesondere bei den Skalen ‚Komplexität' und ‚Neuartigkeit', wo aufgrund fehlenden Fachwissens und mangelnder Erfahrung bzw. Vertrautheit der Laien signifikante Unterschiede zwischen Laien und Experten erwartet worden waren. Auch die Vermutung, nach der Experten das Gesamterlebnis wegen ihrer größeren Erfahrung und damit Vergleichsmöglichkeiten zu anderen Opernhäusern kritischer bewerten könnten, bestätigt sich nicht. Das Ergebnis ist damit dem in Abschnitt 5.2.2 (Untersuchungsfrage 2b) vergleichbar, wo ebenfalls keine signifikanten Unterschiede zwischen Laien und Experten in der Bewertung der künstlerischen Qualität festgestellt werden konnten.

Im Gegensatz zu den Skalenmittelwerten, wo kaum Unterschiede zwischen Laien und Experten bestehen, unterscheiden sich die Gruppen bei Betrachtung der Standardabweichungen der Mittelwerte (siehe Tabelle 26). So ist die Streuung der einzelnen Werte der Experten um den Skalenmittelwert bei den Skalen ‚Gesamturteil', ‚künstlerische Qualität', ‚Emotionen 1' (Genussfaktor/ Ich-Beteiligung) und ‚Emotionen 2' (Empathie/

persönliche Assoziationen) größer, wohingegen die Werte der Laien bei den Skalen ‚Neuartigkeit' und ‚Identifikation' breiter streuen. Die Übereinstimmung der Urteile innerhalb der beiden Gruppen, bei der die Literatur bislang gegenläufige Bilder zeichnet (vgl. Abschnitt 5.2.3), wird daher im nächsten Abschnitt explizit behandelt (Untersuchungsfrage 4b).

5.4.2 Untersuchungsfrage 4b: Übereinstimmung der Gesamturteile innerhalb der beiden Beurteilergruppen

Nachdem festgestellt werden konnte, dass Laien und Experten die Teilaspekte des Gesamterlebnisses eines Opernabends sehr ähnlich bewerten bzw. dass Experten nicht signifikant kritischer urteilen als Laien, wird im Folgenden die Übereinstimmung der Urteile (Interrater-Reliabilität) innerhalb der beiden Gruppen betrachtet (Untersuchungsfrage 4b). Ziel dabei sind Erkenntnisse über das Ausmaß der Subjektivität der Urteile über einen Abend in der Oper. Als Zuverlässigkeitsmaß dienen wiederum Intraklassenkorrelationen (ICC), die die Stärke des Zusammenhangs zwischen Befragten anhand der Korrelation der jeweiligen Werte messen (Wirtz & Caspar, 2002, S. 232, vgl. Abschnitt 5.2.3). Tabelle 27 stellt die Ergebnisse dar.

Tabelle 27: Berechnung der Intraklassenkorrelationen

Skala	ICC_{unjust}		Stichprobengröße	
	Laien	Experten	Laien	Experten
Gesamturteil	.920***	.931***	55	56
künstlerische Qualität	.873***	.893***	57	56
Emotionen 1	.866***	.926***	57	57
Emotionen 2	.746***	.842***	57	56
Komplexität	.536***	.754***	57	56
Neuartigkeit	.864***	.842***	56	56
Identifikation	.920***	.883***	54	54
Aufführungs-saal	.521***	.638***	56	55
Umfeld	.727***	.645***	54	53

Anmerkungen: *** $p \leq .000$. Emotionen 1 = aus den Skalen ‚Genussfaktor' und ‚Ich-Beteiligung' im Rahmen einer explorativen Faktorenanalyse gebildeter Faktor, Emotionen 2 = aus den Skalen ‚Empathie' und ‚persönliche Assoziationen' gebildeter Faktor (vgl. Abschnitt 4.2.3). Da jeweils nur die Befragten berücksichtigt wurden, die sich zu einer Beurteilung in der Lage sahen, d.h. nicht die Antwortmöglichkeiten „habe ich nicht beachtet" bzw. „kann ich nicht beurteilen" wählten, variieren die Stichprobengrößen je nach Konstrukt geringfügig.

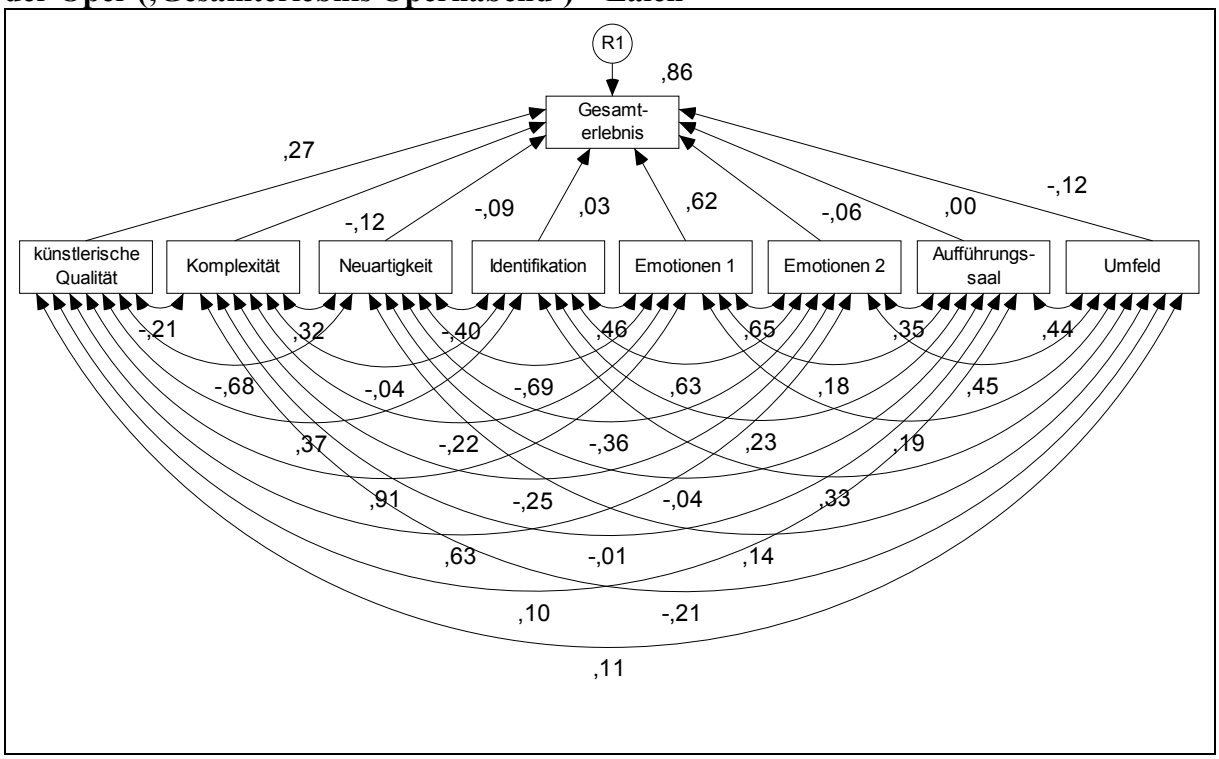
Da Werte von ICC ab 0.7 als gut zu bezeichnen sind (Wirtz & Caspar, 2002, S. 25), kann bei der Mehrheit der Konstrukte von einer hohen Übereinstimmung innerhalb beider Gruppen

ausgegangen werden. Eine Ausnahme bildet das Konstrukt ‚Komplexität', wo bei den Laien intraindividuelle Unterschiede auftreten, d.h. weniger einheitliche Bewertungskriterien bestehen. Ähnliches gilt für die situativen Rahmenbedingungen, wo sowohl innerhalb der Laien- als auch der Expertengruppe keine hohe Übereinstimmung zu vermerken ist und anscheinend unterschiedliche Maßstäbe angelegt werden. Insgesamt jedoch erscheinen Urteile das Gesamterlebnis eines Opernabends betreffend – vergleichbar zur künstlerischen Qualität allein (vgl. Abschnitt 5.2.3) – weit weniger unterschiedlich als ursprünglich vermutet. Ebenfalls kann nicht bestätigt werden, dass die Urteile der Laien weniger konsistent seien als die der Experten; auch hier ergeben sich nur geringe intraindividuelle Unterschiede.

5.4.3 Untersuchungsfrage 4c: Unterschiede zwischen Laien und Experten in der Gewichtung der Teilaspekte

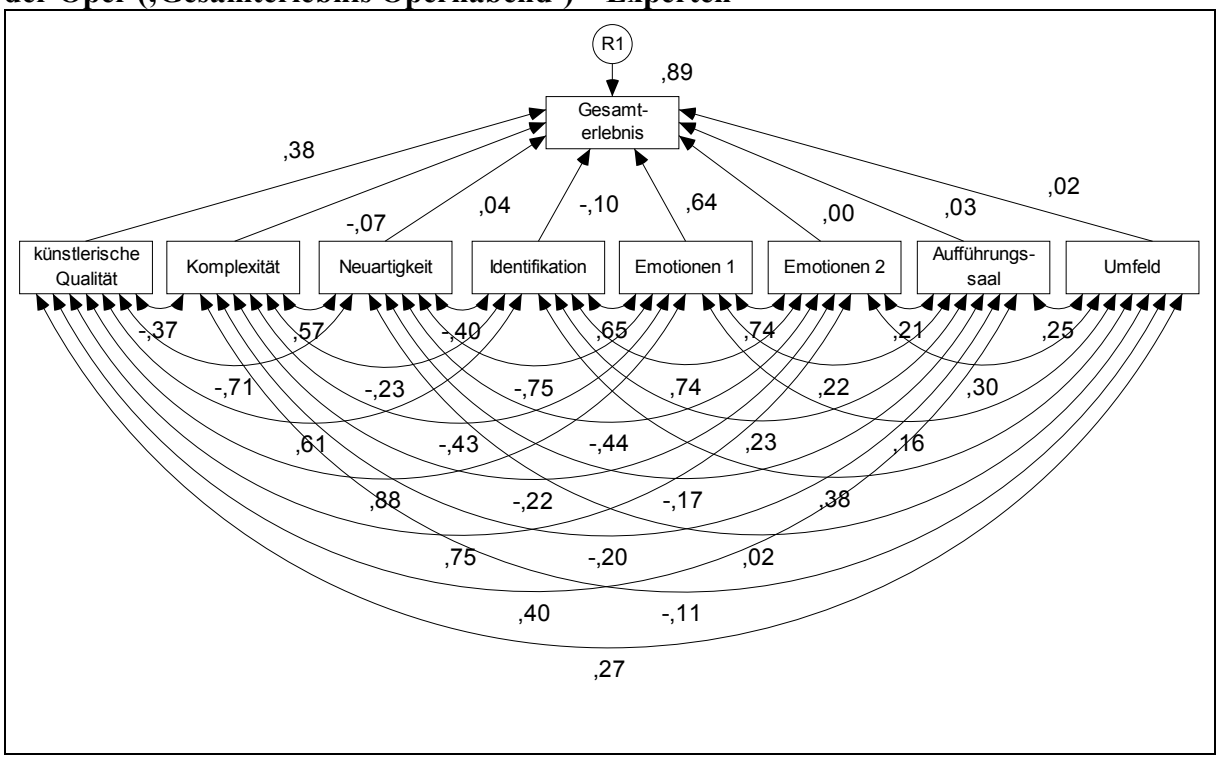
Nach Klärung der Übereinstimmung der Urteile innerhalb der beiden Beurteilergruppen beschäftigt sich Untersuchungsfrage 4c mit den Unterschieden zwischen Laien und Experten hinsichtlich der Gewichtung der Teilaspekte. Die Ergebnisse der separat erfolgten Pfadanalysen können den Abbildungen 12 und 13 entnommen werden.

Abbildung 12: Determinanten des Gesamturteils des Publikums über einen Abend in der Oper (‚Gesamterlebnis Opernabend') – Laien



Anmerkungen zu den Abbildungen 12 und 13: Rechtecke präsentieren manifeste Variablen, Pfeile stellen ihre Zusammenhänge dar. ‚R' präsentiert eine Residualvariable. Werte am rechten oberen Rand der manifesten Variablen stellen die erklärte Gesamtvarianz dar, Werte bei den Pfeilen präsentieren die standardisierten ß-Koeffizienten. Emotionen 1 = aus den Skalen ‚Genussfaktor' und ‚Ich-Beteiligung' im Rahmen einer explorativen Faktorenanalyse gebildeter Faktor, Emotionen 2 = aus den Skalen ‚Empathie' und ‚persönliche Assoziationen' gebildeter Faktor (vgl. Abschnitt 4.2.3).

Abbildung 13: Determinanten des Gesamturteils des Publikums über einen Abend in der Oper (‚Gesamterlebnis Opernabend') – Experten



Wie den Abbildungen 12 bzw. 13 entnommen werden kann, können 86% bzw. 89% der Gesamtvarianz des Urteils über einen Abend in der Oper (‚Gesamterlebnis Opernabend') mit Hilfe der erarbeiteten Determinanten erklärt werden. Bei einem Vergleich der Gewichtung der Teilaspekte des Gesamturteils ergeben sich nur geringe Unterschiede zwischen Laien und Experten. In beiden Beurteilergruppen geht das Konstrukt ‚Genussfaktor/ Ich-Beteiligung' (Emotionen 1) mit der größten Gewichtung in das Urteil über das Gesamterlebnis ein (Laien: ß = .62, Experten ß = .64). An zweiter Stelle steht jeweils das Konstrukt ‚künstlerische Qualität einer Opernaufführung', welches in der Teilstichprobe der Experten einen etwas höheren Stellenwert einnimmt als bei den Laien (Laien: ß = .27, Experten: ß = .38). Dieses Resultat stimmt nicht mit Befunden der musikpsychologischen Literatur überein, wonach Experten Musik eher analytisch, Laien dagegen mehr emotional wahrnehmen (Hargreaves & Colman, 1981; Harrer & Allesch 1982). Inwieweit Ergebnisse aus der Musikpsychologie auf das Musiktheater übertragbar sind oder ob die Untersuchungs-anordnung dieser Arbeit für die sich widersprechenden Ergebnisse verantwortlich zu machen ist, müssen weitere Forschungsprojekte klären. Erwartungsgemäß hingegen sind die geringfügigen Unterschiede, die sich zwischen Laien und Experten hinsichtlich der Konstrukte ‚Komplexität' und ‚Neuartigkeit' ergeben. Diese gehen in der Laiengruppe mit negativen Vorzeichen in das Gesamturteil ein, da die Laien Vertrautes auf der Bühne offenbar schätzen und es daher positiver bewerten.

6. Schlussbetrachtung

6.1 Zusammenfassung der Ergebnisse

Im Rahmen der umfassenden Reformen der öffentlichen Verwaltung, die in Deutschland unter dem Begriff ‚Neues Steuerungsmodell' bekannt wurden, müssen sich auch die öffentlichen Theater einer Evaluation unterziehen. Da jedoch weder Qualitätskonzepte für den For-Profit-Bereich noch solche für Non-Profit-Organisationen für eine Übertragung geeignet sind (Boerner, 2002) und weder die Substanz künstlerischer Qualität im Theater als eindeutig definierbar gilt noch Einigkeit über die Kriterien einer Bewertung herrscht (Heinze, 1994; Lipp, 1994; Hoegl, 1995; Allmann, 1997), ist eine Evaluation äußert schwierig. Mit der vorliegenden Arbeit sollte daher ein Beitrag zur Entwicklung eines Instrumentes geleistet werden, welches sowohl die Entstehung des Publikumsurteils über die künstlerische Qualität einer Opernaufführung als auch über einen gesamten Abend in der Oper (‚Gesamterlebnis Opernabend') erklären kann.

Instrument und Datenerhebung

In einem ersten Schritt wurden daher der von Moser (2005) und Renz (2006) auf Grundlage des analytischen Modells von Boerner (2002) entwickelte Fragebogen zur Erfassung des Publikumsurteils über die künstlerische Qualität im Musiktheater nach Faktoren- und Reliabilitätsanalysen modifiziert sowie zusätzliche Konstrukte zur Erfassung der Determinanten des Gesamturteils des Publikums über einen Abend in der Oper (‚Gesamterlebnis Opernabend') entwickelt. Anschließend kam der Fragebogen bei einer Aufführung der Oper „Die Zauberflöte" von Wolfgang Amadeus Mozart im Anhaltischen Theater Dessau zum Einsatz (n = 120).

Deskriptive Auswertung

Um Erkenntnisse darüber zu erhalten, welche Aspekte einer Opernaufführung von Zuschauern nicht beachtet werden bzw. nicht beurteilt werden können, wurde die Häufigkeit des Auftretens der Antwortmöglichkeiten „habe ich nicht beachtet" bzw. „kann ich nicht beurteilen" herangezogen. Diese wurden den Teilnehmern zur Beantwortung der Items des Fragebogens zur Verfügung gestellt, um Informationen darüber zu erhalten, inwieweit das analytische Modell Boerners (2002) mit der Struktur der empirisch wahrgenommenen künstlerischen Qualität einer Opernaufführung übereinstimmt. Zwar nahm die Anzahl der Items, die von einem Teil der Befragten nicht beantwortet wurde, im Vergleich zu den

Untersuchungen von Renz (2006) und Boerner et al. (in Druck) insgesamt erheblich ab, jedoch konnte neben einer hohen Übereinstimmung des analytischen Modells mit den empirischen Daten anhand der Häufigkeit der Antwortkategorie „kann ich nicht beurteilen“ festgestellt werden, dass im Fragebogen mitunter nach Items gefragt wurde, die zwar im Analyseraster der Zuschauer, jedoch außerhalb der selbsteingeschätzten Urteilskompetenz lagen. Bei einer erneuten Modifikation des Fragebogens werden diese Items daher eine spezielle Aufmerksamkeit erfordern, um dem Ziel, nur noch Konstrukte abzufragen, die vom Publikum für die Bewertung der künstlerischen Qualität einer Opernaufführung als relevant betrachtet werden, sowie ausschließlich solche Items zu verwenden, die ein durchschnittliches Opernpublikum kompetent zu beantworten in der Lage ist, näher zu kommen.

Reliabilitäts- und Faktorenanalysen

Anschließende Reliabilitäts- und Faktorenanalysen dienten der Validierung des von Moser (2005) und Renz (2006) zur Erfassung des Publikumsurteils über die künstlerische Qualität im Musiktheater entwickelten und in der vorliegenden Arbeit modifizierten Fragebogens. Als Ergebnis kann festgehalten werden, dass die Güte des Fragebogens dieser Arbeit die des Fragebogens von Renz (2006) deutlich übersteigt und als sehr gut bezeichnet werden kann. Nach erfolgter Validierung hätte als nächster Schritt eine weitere Datenauswertung erfolgen müssen, um die Untersuchungsfragen zu beantworten. Da dies im Rahmen der vorliegenden Arbeit aus Zeitgründen jedoch nicht möglich war, wurden sie anhand der bereits erhobenen Daten zu beantworten versucht.

Untersuchungsfrage 1

Als Ergebnis der Untersuchungsfrage 1a – welche musikalischen und szenischen Aspekte einer Opernaufführung das Publikum bei seinem Urteil beachtet und diese dann in welchem Differenzierungsgrad unterscheidet – kann festgehalten werden, dass die ‚Stimmigkeit zwischen der musikalischen und der szenischen Dimension’, die ‚Stimmigkeit innerhalb der musikalischen Dimension’, die ‚Stimmigkeit innerhalb der szenischen Dimension’ sowie die physisch unterscheidbaren Konstrukte ‚Orchester/ Dirigent’, ‚Chor’, ‚Solist’ und ‚Ausstattung’ vom Publikum getrennt wahrgenommen werden. Im Gegensatz zur Annahme des analytischen Modell Boerners (2002) differenzieren die Zuschauer jedoch nicht zwischen den verschiedenen Funktionen der Akteure, lediglich beim Chor wird zwischen seiner musikalischen und seiner szenischen Leistung unterschieden. Zur Bestimmung der Gewichtung, mit der die unterschiedlichen Teilaspekte in das Gesamturteil der Zuschauer

über die künstlerische Qualität einer Opernaufführung eingehen (Untersuchungsfrage 1b) wurden Pfadanalysen durchgeführt. Das größte Gewicht auf das Gesamturteil hat demnach die ‚Stimmigkeit zwischen der musikalischen und der szenischen Dimension'. Es folgen die ‚Stimmigkeit innerhalb der szenischen Dimension', die ‚Stimmigkeit innerhalb der musikalischen Dimension' und das Konstrukt ‚Szene/ Ausstattung'. Nur eine verhältnismäßig geringe Rolle spielen an dieser Stelle die Solisten, das Orchester sowie der Chor.

Untersuchungsfrage 2

Da sich das Publikum in Hinblick auf musikalisches Fachwissen und Opernerfahrung unterscheidet, woraus Konsequenzen für die Urteilsbildung erwartet werden, wurde in Untersuchungsfrage 2 zwischen Laien und Experten im Publikum unterschieden. Untersuchungsfrage 2a beschäftigte sich mit Unterschieden zwischen den beiden Beurteilergruppen in Bezug auf die Differenziertheit und die Gewichtung der Teilaspekte des Urteils über die künstlerische Qualität. Separate Faktoren- und Pfadanalysen ergaben jedoch nur geringe Unterschiede. Auch die Betrachtung der Bewertung der Qualität einer Opernaufführung getrennt nach Laien und Experten (Untersuchungsfrage 2b) zeigt nur geringe Differenzen. Lediglich bei den Konstrukten ‚musikalische Leistung und Zusammenspiel innerhalb des Orchesters' sowie ‚musikalische Leistung des Chores' bestanden signifikante Unterschiede zwischen Laien und Experten. Es wird vermutet, dass das speziell zur Beurteilung der musikalischen Dimension notwenige Fachwissen und die Vergleiche ermöglichende Erfahrung nur den Experten zur Verfügung stand, die daher kritischer urteilten. Untersuchungsfrage 2c widmete sich der Höhe der Übereinstimmung der Urteile innerhalb der beiden Beurteilergruppen. Mit Hilfe der Berechnung von Intraklassenkorrelationen wurde bei allen Skalen ein hoher Grad an Übereinstimmung der Urteile innerhalb der jeweiligen Gruppen festgestellt; die künstlerischer Qualität in der Oper scheint daher sowohl in der Teilstichprobe der Laien als auch bei den Experten anhand einheitlicher Kriterien beurteilt zu werden.

Untersuchungsfrage 3

Im zweiten Teil der vorliegenden Arbeit standen die Determinanten des Gesamturteils des Publikums über einen Abend in der Oper (‚Gesamterlebnis Opernabend') im Fokus. Da sowohl die Literatur zur Musikrezeption als auch die zum Sprechtheater gezeigt hatten, dass neben der Bewertung der künstlerischen Qualität weitere Aspekte in das Gesamturteil des Publikums über einen Abend im Konzert bzw. im Sprechtheater einfließen, und in dieser

Arbeit ähnliche Effekte auch im Musiktheater vermutet wurden, wurden bei der Konstruktion des Fragebogens zusätzliche Aspekte erhoben. Bei der Durchführung einer Pfadanalyse zur Beantwortung der Untersuchungsfrage 3 – welche Determinanten das Gesamturteils des Publikums über einen Abend in der Oper (‚Gesamterlebnis Opernabend') bestimmen – erwies sich das Konstrukt ‚Genussfaktor/ Ich-Beteilung', welches zur emotionalen Wirkung der Aufführung auf den Zuschauer gehört, als die wichtigste Determinante. An zweiter Stelle stand die künstlerische Qualität, wohingegen Aspekte wie die Identifikation mit Charakteren auf der Bühne oder die Neuartigkeit der Inszenierung keine Rolle für das Publikumsurteil über den Abend in der Oper spielten. Das große Gewicht der emotionalen Wirkung der Aufführung auf das Publikumsurteil über einen Abend in der Oper überrascht insofern, als dass die Literatur bislang auf die Bewertung der künstlerischen Qualität fokussiert war (Thompson, 2007). Die Erkenntnis über die entscheidende Rolle der Emotionen im Musiktheater kann daher als innovatives Ergebnis der vorliegenden Arbeit gesehen werden.

Untersuchungsfrage 4

Auch bei der Beurteilung des Gesamturteils des Publikums über einen Abend in der Oper (‚Gesamterlebnis Opernabend') interessierten Unterschiede zwischen Laien und Experten; dieser Differenz widmete sich Untersuchungsfrage 4. So ergab die separate Untersuchung der Bewertung der Teilaspekte des Gesamterlebnisses (Untersuchungsfrage 4a) keine signifikanten Unterschiede zwischen beiden Beurteilergruppen. Zur Untersuchung der Höhe der Übereinstimmung der Urteile innerhalb der Laien- und der Expertengruppe (Untersuchungsfrage 4b) mit dem Ziel der Bestimmung des Ausmaßes der Individualität der Urteile über einen Abend in der Oper wurden wiederum die Intraklassenkorrelationen berechnet. Mit Ausnahme der Konstrukte ‚Komplexität der Aufführung' und ‚situative Rahmenbedingungen', wozu bspw. die Akustik des Saales und die Atmosphäre im Theater gehören, kann bei allen Konstrukten von einer hohen Übereinstimmung, d.h. einheitlichen Bewertungskriterien, ausgegangen werden. Abschließend setzte sich Untersuchungsfrage 4c mit den Unterschieden zwischen Laien und Experten in Bezug auf die Gewichtung der Teilaspekte des Gesamturteils über einen Abend in der Oper (‚Gesamterlebnis Opernabend') auseinander. Wiederum ergaben sich nur geringe Differenzen: Sowohl bei den Laien als auch bei den Experten ging das Konstrukt ‚Genussfaktor/ Ich-Beteiligung' mit der größten Gewichtung in das Urteil über das Gesamterlebnis ein, an zweiter Stelle steht jeweils das Konstrukt ‚künstlerische Qualität einer Opernaufführung'. Das Ergebnis stimmt demnach nicht mit der musikpsychologischen Literatur überein, wonach Experten Musik eher

analytisch, Laien dagegen mehr emotional wahrnehmen (Hargreaves & Colman, 1981; Harrer & Allesch 1982).

6.2 Diskussion der Ergebnisse

Die vorliegende Studie untersuchte anhand einer Befragung des Publikums einer Aufführung der Oper „Die Zauberflöte“ die Wahrnehmung und Beurteilung der künstlerischen Qualität einer Opernaufführung durch das Publikum sowie die Determinanten seines Gesamturteils über einen Abend in der Oper. Aufgrund des begrenzten Rahmens der Arbeit konnten nicht sämtliche Variablen erhoben werden, von denen ein Einfluss auf das Publikumsurteil in der Oper vermutet wurde. So zeigte die Literaturanalyse, dass Urteile über eine Aufführung immer im Kontext anderer Menschen gefällt werden und gerade in „nicht eindeutigen Wahrnehmungssituationen zuvor abgegebene Urteile anderer das eigene Urteil beeinflussen“ können (Behne, 1997, S. 1006). Auch die Individualgeschichte eines Zuschauers sowie seine Persönlichkeit können einen Einfluss auf das Urteil über die Aufführung bzw. den gesamten Abend ausüben (Adorno, 1973; Rawlings et al., 1998; McNamara & Ballard, 1999; Weaver, 2000). Daher ist anzunehmen, dass nicht erhobene Variablen dieser Art für einen Teil der nicht aufgeklärten Varianz des Gesamturteils über einen Abend in der Oper verantwortlich sind.

Als größte Schwäche dieser Untersuchung muss jedoch die Tatsache gesehen werden, dass lediglich das Publikum *einer* Aufführung *einer* Produktion an *einem* Theater befragt wurde. Aussagen über eine Aufführungs-, Produktions- bzw. Werkabhängigkeit des Publikumsurteils waren daher nicht möglich. Da sich sowohl bei der Untersuchung von Moser (2005) als auch bei der von Renz (2006) Produktions- bzw. Werkabhängigkeiten ergaben, bietet es sich für weitere Forschungsprojekte an, das Publikum unterschiedlicher Aufführungen, unterschiedlicher Produktionen sowie unterschiedlicher Opernhäuser zu befragen. Auch sollten Werke unterschiedlicher Epochen und Inszenierungstypen gewählt werden, um einen Vergleich der Beurteilungen älterer und neuerer Werke bzw. Inszenierungen zu ermöglichen.

Des Weiteren ist die geringe Größe der Stichprobe der vorliegenden Studie zu kritisieren (n = 120), die eine Untersuchung der Gesamtstruktur des Publikumsurteils anhand von Strukturgleichungsmodellen nicht erlaubte. So war in dieser Arbeit nur die Untersuchung von Teilstrukturen des Urteils über die künstlerische Qualität im Rahmen von Strukturgleichungsmodellen möglich; die Gesamtstruktur des Urteils konnte mit Pfadanalysen

untersucht werden, die aufgrund der niedrigeren Anzahl zu schätzender Parameter einen geringeren Stichprobenumfang benötigen. Da Strukturgleichungsmodelle jedoch zusätzlich die Berücksichtigung latenter Variablen und damit genauere Aussagen erlauben, wären weitere Untersuchungen der Publikumsurteile mit großen Stichproben wünschenswert.

Ebenfalls kritisch gesehen werden muss die Tatsache, dass die Auswahl der Befragten keine Zufallsstichprobe darstellte. Diese ist ausschließlich bei Auswahlen gegeben, „bei denen ein Zufallsprozess über die Aufnahme eines Elementes in die Stichprobe entscheidet“ (Schnell et al., 2005, S. 273). Aus zwei Gründen war dies in der vorliegenden Untersuchung jedoch nicht gegeben: Zum einen konnten aufgrund des schmalen Zeitfensters, das zum Austeilen des Fragebogens bestand, nicht sämtliche Zuschauer der Aufführung um ihre Teilnahme an der Befragung gebeten werden. Es ist daher zu befürchten, dass sympathisch bzw. zugänglich wirkende Personen überproportional oft ausgewählt wurden (Reuband, 2005, S. 124). Schwerer allerdings wiegt die Tatsache, dass von den angesprochenen Personen nur gut 30% der Teilnahme an der Befragung zustimmten. Zu befürchten ist daher, dass die Ausfälle systematisch erfolgten, da generell Menschen mit höherem Bildungsniveau und stark am Untersuchungsgegenstand interessierte Personen eher an Befragungen teilnehmen (sog. Selbstrekrutierung, Schnell et al., 2005, S. 359). Da durch einen Vergleich der erhobenen demografischen Merkmale der Befragungsteilnehmer mit weiteren Untersuchungen über Zuschauer von Opernaufführungen zumindest ausgeschlossen werden konnte, dass deutliche Unterschiede bzgl. der Alters- bzw. Geschlechtsverteilung bestanden und die Ergebnisse denen der Studien von Moser (2005), Renz (2006) und Boerner et al. (in Druck) ähnlich sind, muss jedoch nicht von erheblichen Stichprobenverzerrungen ausgegangen werden.

Ausgehend von den in Abschnitt 6.1 zusammenfassten Ergebnissen sollte der Fragebogen vor einem erneuten Einsatz wiederum modifiziert werden. So ist aufgrund des unverzichtbaren Informationsgehaltes der Antwortkategorien „habe ich nicht beachtet“ bzw. „kann ich nicht beurteilen“ in Erwägung zu ziehen, diese den Befragten bei sämtlichen Items des Fragebogens zur Verfügung zu stellen. Darüber hinaus sollten die Items, bei denen sowohl die Laien als auch die Experten Schwierigkeiten bei der Beantwortung hatten, kritisch diskutiert bzw. überarbeitet werden. Darüber hinaus wurde bei einigen Konstrukten festgestellt, dass das Publikum zu einer Differenzierung in erwartetem Ausmaß nicht in der Lage ist. So erscheint bei einer weiteren Modifikation des Fragebogens bspw. eine Zusammenlegung der Ausstattungsaspekte sinnvoll. Auch die Items zur Erfassung der emotionalen Wirkung einer

Aufführung auf den Zuschauer bedürfen einer Überarbeitung, da die Ergebnisse der explorativen Faktorenanalyse nicht den erwarteten entsprachen.

Anhang A: Fragebogen zur Bewertung von Aufführungen in der Oper

Anmerkungen: Zum besseren Verständnis wurden die Kennziffern der Items (z.B. mGE1) sowie die Abkürzungen der jeweiligen Konstrukte (z.B. fit3_Musik) nachträglich eingefügt, da in der vorliegenden Arbeit die Abkürzungen verwendet werden. Items, die im Vergleich zur Version von Renz (2006, siehe Anhang B) neu konstruiert bzw. umformuliert wurden, sind grau hinterlegt.

Johanna Jobst (Diplomandin)
Lehrstuhl für Management, insbesondere Strategie und Führung
Fachbereich Politik- und Verwaltungswissenschaft
Universität Konstanz

Liebe Besucherin, lieber Besucher,

um künftig noch gezielter auf **Ihre Anregungen und Wünsche** einzugehen, führt die Universität Konstanz in Zusammenarbeit mit dem Anhaltischen Theater Dessau heute eine Besucherbefragung durch. **Dafür bitten wir um Ihre Mithilfe!**

Das Ausfüllen des Fragebogens wird **10 bis 15 Minuten** in Anspruch nehmen. Bitte füllen Sie den Fragebogen **persönlich** aus und geben Sie **Ihre ganz persönliche Meinung** ab. Bitte antworten Sie zügig und ohne lange zu überlegen; es gibt keine „richtigen“ oder „falschen“ Antworten.

Der Fragebogen enthält Fragen zu den einzelnen Aspekten der Aufführung „Die Zauberflöte“. Einige Fragen sind ähnlich formuliert – dies ist für die Auswertung notwendig und daher gewollt. Bitte lesen Sie sich die Instruktionen sorgfältig durch. Bitte setzen Sie **nur ein Kreuz pro Frage** und markieren Sie Ihr Kreuz bei der Antwort, die **Ihren persönlichen Eindruck** von der Aufführung am besten beschreibt. Bitte beachten Sie, dass die Seiten **beidseitig bedruckt** sind. Beantworten Sie möglichst jede Frage.

Ideal für die Befragung wäre es, wenn Sie den Fragebogen **nach der Aufführung** beantworten und in die bereitgestellten Kästen am Ausgang geben. Wenn dies für Sie nicht möglich sein sollte, haben Sie die Möglichkeit, den Fragebogen bereits **in der Pause** auszufüllen. Es besteht ebenfalls die Möglichkeit, den Fragebogen **zu Hause** auszufüllen und ihn an uns zu senden. Einen vorbereiteten Freiumschlag erhalten Sie am Ausgang.

Selbstverständlich werden Ihre Aussagen absolut **vertraulich** behandelt. Aus Ihren Angaben sind keinerlei Rückschlüsse auf Ihre Person möglich.

Als kleines Dankeschön für Ihre Unterstützung verlosen wir unter allen Teilnehmern **5 x zwei Theaterkarten** des Anhaltischen Theaters Dessau.

Vielen Dank für Ihre Mitarbeit!

A. Gesamteindruck

Kennziffer des Items	***Wie hat Ihnen der heutige Abend in der Oper insgesamt gefallen?***	**Trifft voll und ganz zu**	**Trifft eher zu**	**Teils, teils**	**Trifft eher nicht zu**	**Trifft über-haupt nicht zu**
mGE1	Meine Erwartungen an den Abend wurden vollständig erfüllt.	❐	❐	❐	❐	❐
mGE2	Der Abend war für mich insgesamt ein tolles Erlebnis.	❐	❐	❐	❐	❐
mrGE3	Der Abend hat mich sehr enttäuscht.	❐	❐	❐	❐	❐
mGE4	Dieser Opernabend wird mir lange positiv in Erinnerung bleiben.	❐	❐	❐	❐	❐
mGE5	Veranlasst von diesem Abend habe ich mir vorgenommen, öfter eine Opernaufführung zu besuchen.	❐	❐	❐	❐	❐
mGQ1	Die künstlerische Qualität der Aufführung hat mich insgesamt sehr überzeugt.	❐	❐	❐	❐	❐
mGQ2	Ich bin der Meinung, dass die Aufführung eine künstlerische Spitzenleistung war.	❐	❐	❐	❐	❐
urGU1	Es wäre kein Verlust, wenn diese Produktion schon bald aus dem Spielplan genommen würde.	❐	❐	❐	❐	❐
vNG2	Ich würde erwarten, dass die Aufführung von der Kritik sehr gut aufgenommen und besprochen wird.	❐	❐	❐	❐	❐

Kennziffer des Items	***Wie hat Ihnen die heutige Aufführung von der Inszenierung her gefallen?***	**Trifft voll und ganz zu**	**Trifft eher zu**	**Teils, teils**	**Trifft eher nicht zu**	**Trifft über-haupt nicht zu**
rKA1_k	Auf der Bühne war oft soviel los, dass man nicht wusste, wo man eigentlich hinschauen sollte.	❐	❐	❐	❐	❐

Kennziffer des Items	***(Fortsetzung)*** ***Wie hat Ihnen die heutige Aufführung <u>von der Inszenierung her</u> gefallen?***	**Trifft voll und ganz zu**	**Trifft eher zu**	**Teils, teils**	**Trifft eher nicht zu**	**Trifft über-haupt nicht zu**
rKA2_k	Es war oft schwierig, den genauen Fortgang der Handlung zu verfolgen.	❐	❐	❐	❐	❐
rKA3_k	Die zu häufige Bewegung der Darsteller im Raum lenkte von der eigentlichen Handlung ab.	❐	❐	❐	❐	❐
rKA9_v	Die Art der Aufführung hat mich ziemlich irritiert.	❐	❐	❐	❐	❐
ruGA6_v	Die Inszenierung war mir insgesamt zu neuartig und modern.	❐	❐	❐	❐	❐
vrKA7_v	Die Inszenierung verwendete viele verwirrende Stilmittel.	❐	❐	❐	❐	❐
I1_g	Das Gesamtkonzept der Inszenierung war für mich gut erkennbar.	❐	❐	❐	❐	❐
rI3_g	Die Inszenierung ergab in meinen Augen keine schlüssige Botschaft.	❐	❐	❐	❐	❐
I5_g	Der Grundgedanke dieser Inszenierung hat mich absolut überzeugt.	❐	❐	❐	❐	❐
mII1_i	Das Verhalten des Papageno bzw. der Papagena ist für mich vorbildlich.	❐	❐	❐	❐	❐
mII2_i	Ich konnte mich mit der Rolle des Papageno bzw. der Papagena sehr gut identifizieren.	❐	❐	❐	❐	❐
mII3_i	Papageno bzw. Papagena stehen für Werte, die mir persönlich sehr wichtig sind.	❐	❐	❐	❐	❐

Kennziffer des Items	***Bitte beantworten Sie hier, wie Sie die heutige Aufführung <u>insgesamt empfunden</u> haben.***	**Trifft voll und ganz zu**	**Trifft eher zu**	**Teils, teils**	**Trifft eher nicht zu**	**Trifft überhaupt nicht zu**
mEE1	Ich konnte die Gefühle der Darsteller sehr gut nachempfinden.	❐	❐	❐	❐	❐
mEE2	Während der Aufführung fieberte ich mit dem Schicksal der Darsteller mit.	❐	❐	❐	❐	❐
mEE3	Ich konnte mit den Darstellern mitfühlen, als ich in dieser Aufführung saß.	❐	❐	❐	❐	❐

Kennziffer des Items	***(Fortsetzung)*** ***Bitte beantworten Sie hier, wie Sie die heutige Aufführung <u>insgesamt empfunden</u> haben.***	Trifft voll und ganz zu	Trifft eher zu	Teils, teils	Trifft eher nicht zu	Trifft überhaupt nicht zu
mEA1	Die Aufführung weckte positive Erinnerungen in mir.	❐	❐	❐	❐	❐
mEA2	Während der Aufführung war ich zeitweise traurig, weil negative Erinnerungen in mir geweckt wurden.	❐	❐	❐	❐	❐
mEA3	Die Aufführung war in der Lage, mich an Ereignisse und Situationen aus meinem eigenen Leben zu erinnern.	❐	❐	❐	❐	❐
mEA4	In dieser Aufführung wurde ich immer wieder an meine Alltagsprobleme erinnert.	❐	❐	❐	❐	❐
mEP1	Ich genoss die Aufführung und fühlte mich wohl.	❐	❐	❐	❐	❐
mEP2	Während der Aufführung war ich froh und entspannt.	❐	❐	❐	❐	❐
mEI1	Während der Aufführung verging der Zeit wie im Flug.	❐	❐	❐	❐	❐
vKA4_EI	Ich konnte alles um mich herum vergessen, als ich in dieser Aufführung saß.	❐	❐	❐	❐	❐
mEI3	Ich war emotional „Feuer und Flamme“, als ich in dieser Aufführung saß.	❐	❐	❐	❐	❐
mES1	Während der Aufführung gab es immer wieder spezielle Momente, in denen ich besonders emotional auf das Geschehen auf der Bühne reagierte (z.B. Freude, Trauer, Spannung).	❐	❐	❐	❐	❐
mES2	Während der Aufführung reagierte ich teilweise auch physisch (z.B. Herzklopfen, feuchte Hände, Frösteln, Tränen).	❐	❐	❐	❐	❐
mES3	Zwischendurch war ich ziemlich aufgewühlt.	❐	❐	❐	❐	❐
rNG7	Die Aufführung hat mich insgesamt ziemlich kalt gelassen.	❐	❐	❐	❐	❐
vNG6	Die Aufführung hat mich gefühlsmäßig sehr berührt.	❐	❐	❐	❐	❐
mrEM3	Ich fühlte mich von der Aufführung emotional nicht angesprochen.	❐	❐	❐	❐	❐

B. Bewertung der Einzelaspekte

In diesem Teil des Fragebogens interessiert uns, wie Sie ***das Orchester bzw. den Dirigenten, den Solisten in der Rolle des Papageno und den Chor in der heutigen Aufführung jeweils für sich genommen*** *erlebt haben. Bitte bewerten Sie zunächst die musikalische Leistung und anschließend die szenische/darstellerische Leistung des Ensembles.*

Bitte beachten Sie, dass Sie im folgenden Teil auch die Antwortmöglichkeiten „habe ich nicht beachtet" und „kann ich nicht beurteilen" zur Verfügung haben. Bitte kreuzen Sie diese nur an, wenn Sie die Frage ***wirklich nicht*** *beantworten können.*

Musikalische Leistung des Ensembles

Kennziffer des Items	*Wie hat Ihnen das Orchester bzw. der Dirigent gefallen?*	Trifft voll und ganz zu	Trifft eher zu	Teils, teils	Trifft eher nicht zu	Trifft über-haupt nicht zu	Habe ich nicht beach-tet	Kann ich nicht beur-teilen
mMO13	Das Orchester erschien mir sehr gut vorbereitet.	❐	❐	❐	❐	❐	❐	❐
vMO2	Unterschiede in der Lautstärke wurden vom Orchester sehr gut herausgearbeitet.	❐	❐	❐	❐	❐	❐	❐
vuMO3	Dem Orchester ist es sehr gut gelungen, den ganzen Abend über die musikalische Spannung zu halten.	❐	❐	❐	❐	❐	❐	❐
vMO5	Das Orchester spielte mit sehr großer Spannung und Intensität.	❐	❐	❐	❐	❐	❐	❐
MO10	Das Orchester spielte sehr lebendig.	❐	❐	❐	❐	❐	❐	❐
vfit3O11	Die verschienen Instrumentengruppen des Orchesters (Streicher, Bläser, Schlagwerk) waren sehr gut aufeinander abgestimmt.	❐	❐	❐	❐	❐	❐	❐
fit3O5	Das Orchester spielte „wie aus einem Guss".	❐	❐	❐	❐	❐	❐	❐
MD1	Der Dirigent hatte das Orchester spürbar gut „im Griff".	❐	❐	❐	❐	❐	❐	❐
rMD4	Der Dirigent schien kein Konzept zu haben: Der Aufführung fehlte die musikalische Linie über die verschiedenen Nummern hinweg.	❐	❐	❐	❐	❐	❐	❐
vMD7	Der Dirigent verstand es sehr gut, die Balance zwischen Orchestergraben und Sängern zu halten.	❐	❐	❐	❐	❐	❐	❐

Kennziffer des Items	*(Fortsetzung) Wie hat Ihnen das Orchester bzw. der Dirigent gefallen?*	Trifft voll und ganz zu	Trifft eher zu	Teils, teils	Trifft eher nicht zu	Trifft über-haupt nicht zu	Habe ich nicht beach-tet	Kann ich nicht beur-teilen
MD13	Der Dirigent zeigte sich dem Werk in jeder Hinsicht gewachsen.	❐	❐	❐	❐	❐	❐	❐

Kennziffer des Items	*Wie hat Ihnen der Solist in der Rolle des Papageno gefallen?*	Trifft voll und ganz zu	Trifft eher zu	Teils, teils	Trifft eher nicht zu	Trifft über-haupt nicht zu	Habe ich nicht beach-tet	Kann ich nicht beur-teilen
MS1	Papageno sang bei seinen Soli mit einem hohen Nuancenreichtum.	❐	❐	❐	❐	❐	❐	❐
vrMS2	Die Stimme des Papageno besaß sehr wenig klangliche Schönheit.	❐	❐	❐	❐	❐	❐	❐
MS5	Die Stimme des Papageno war ideal für die Anforderungen der Partie.	❐	❐	❐	❐	❐	❐	❐
vMS6	Papageno brachte die Unterschiede in der Lautstärke seiner Partie sehr gut zur Geltung.	❐	❐	❐	❐	❐	❐	❐
vuMS7	Papageno war seiner Partie musikalisch insgesamt sehr gut gewachsen.	❐	❐	❐	❐	❐	❐	❐
MS10	Papageno wurde dem Anspruch der Rolle musikalisch gerecht.	❐	❐	❐	❐	❐	❐	❐

Kennziffer des Items	*Wie hat Ihnen der Chor gefallen?*	Trifft voll und ganz zu	Trifft eher zu	Teils, teils	Trifft eher nicht zu	Trifft über-haupt nicht zu	Habe ich nicht beach-tet	Kann ich nicht beur-teilen
MC2	Der Chor brachte Unterschiede in der Lautstärke sehr gut zur Geltung.	❐	❐	❐	❐	❐	❐	❐
mMC10	Die musikalische Leistung des Chores hat mir sehr gefallen.	❐	❐	❐	❐	❐	❐	❐
MC8	Der Chor zeigte sich dem Stück in jeder Hinsicht gewachsen.	❐	❐	❐	❐	❐	❐	❐
MC7	Der Chor überzeugte durch seine klangliche Brillanz.	❐	❐	❐	❐	❐	❐	❐

Im Folgenden interessiert uns Ihre Einschätzung der szenischen/darstellerischen Leistungen.

Kennziffer des Items	***Wie hat Ihnen der Solist in der Rolle des Papageno gefallen?***	Trifft voll und ganz zu	Trifft eher zu	Teils, teils	Trifft eher nicht zu	Trifft über-haupt nicht zu	Habe ich nicht beach-tet	Kann ich nicht beur-teilen
SS1	Papageno spielte die Figur eindringlich heraus.	❐	❐	❐	❐	❐	❐	❐
vSS4	Papageno lebte die Rolle so, dass ich mich sehr gut mit ihr identifizieren konnte.	❐	❐	❐	❐	❐	❐	❐
SS5	Papageno verkörperte die Charaktereigenschaften der Figur sehr überzeugend.	❐	❐	❐	❐	❐	❐	❐
vSS8	Papagenos Kostüm passte sehr gut zu seiner Rolle.	❐	❐	❐	❐	❐	❐	❐
rSS10	Mimik, Gestik und Bewegungen von Papageno passten oft nicht zum szenischen Geschehen.	❐	❐	❐	❐	❐	❐	❐

Kennziffer des Items	***Wie hat Ihnen der Chor gefallen?***	Trifft voll und ganz zu	Trifft eher zu	Teils, teils	Trifft eher nicht zu	Trifft über-haupt nicht zu	Habe ich nicht beach-tet	Kann ich nicht beur-teilen
vuSC1	Das äußere Erscheinungsbild der Chorsänger (Kostüm und Maske) passte sehr gut zur Rolle des Chores in dieser Aufführung.	❐	❐	❐	❐	❐	❐	❐
vSC2	Ich konnte dem Chor durch sein Auftreten und Aussehen seine Rolle voll abnehmen.	❐	❐	❐	❐	❐	❐	❐
vSC3	Das äußere Erscheinungsbild der Chorsänger (Kostüm und Maske) war ästhetisch sehr befriedigend.	❐	❐	❐	❐	❐	❐	❐
vuSC4	Die Anordnung und Bewegung des Chores im Raum passte sehr gut zur Rolle des Chores.	❐	❐	❐	❐	❐	❐	❐
SC6	Schauspielerisch fand ich den Chor insgesamt sehr überzeugend.	❐	❐	❐	❐	❐	❐	❐

Kennziffer des Items	***Wie hat Ihnen die Ausstattung gefallen?***	Trifft voll und ganz zu	Trifft eher zu	Teils, teils	Trifft eher nicht zu	Trifft überhaupt nicht zu
vSA13	Das Bühnenbild hat mir insgesamt sehr gut gefallen.	❐	❐	❐	❐	❐

Kennziffer des Items	*(Fortsetzung)* *Wie hat Ihnen die Ausstattung gefallen?*	Trifft voll und ganz zu	Trifft eher zu	Teils, teils	Trifft eher nicht zu	Trifft überhaupt nicht zu
mSA16	Ich fand das Bühnenbild sehr beeindruckend.	❐	❐	❐	❐	❐
vSA15	Die Kostüme haben mir insgesamt sehr gut gefallen.	❐	❐	❐	❐	❐
mSA17	Die Kostüme waren sehr wirkungsvoll.	❐	❐	❐	❐	❐

C. Bewertung des Zusammenspiels

Kennziffer des Items	*Bitte beurteilen Sie das Zusammenspiel innerhalb des Chores.*	Trifft voll und ganz zu	Trifft eher zu	Teils, teils	Trifft eher nicht zu	Trifft über-haupt nicht zu	Habe ich nicht beach-tet	Kann ich nicht beur-teilen
fit3C1	Der Chor wirkte in sich klanglich sehr homogen.	❐	❐	❐	❐	❐	❐	❐
vufit3C2	Die Chorstimmen waren in Bezug auf das Tempo sehr gut aufeinander abgestimmt.	❐	❐	❐	❐	❐	❐	❐
vfit3C3	Der Chor wirkte klanglich ausgewogen, so dass nie Stimmen einzelner Personen deutlich herauszuhören waren.	❐	❐	❐	❐	❐	❐	❐

Kennziffer des Items	*Bitte beurteilen Sie das Zusammenspiel aller Solisten untereinander.* *Bitte beziehen Sie Ihre Antworten jetzt auf das **gesamte Ensemble**, nicht nur auf die Rolle des Papageno.*	Trifft voll und ganz zu	Trifft eher zu	Teils, teils	Trifft eher nicht zu	Trifft über-haupt nicht zu	Habe ich nicht beach-tet	Kann ich nicht beur-teilen
vfit3S1	Die Solisten waren klanglich untereinander sehr gut abgestimmt.	❐	❐	❐	❐	❐	❐	❐
rfit3S3	Es kam häufig vor, dass die Solisten mehr „gegeneinander“ als miteinander musizierten.	❐	❐	❐	❐	❐	❐	❐
ufit3S5	Das Solistenensemble wirkte musikalisch untereinander ausgesprochen harmonisch.	❐	❐	❐	❐	❐	❐	❐
urfit3S6	Man hatte den Eindruck, dass sich die Solisten gegenseitig „an die Wand spielen/singen.“	❐	❐	❐	❐	❐	❐	❐

Kennziffer des Items	***Bitte beurteilen Sie hier das Zusammenspiel von Orchester, Solisten und Chor.***	Trifft voll und ganz zu	Trifft eher zu	Teils, teils	Trifft eher nicht zu	Trifft über-haupt nicht zu	Habe ich nicht beach-tet	Kann ich nicht beur-teilen
fit2_2	Die Solisten hatten genügend Strahlkraft und Volumen, um gegen das Orchester „anzukommen“.	❐	❐	❐	❐	❐	❐	❐
vfit2_3	Orchester, Chor und Solisten waren in Bezug auf die Lautstärke sehr gut aufeinander abgestimmt.	❐	❐	❐	❐	❐	❐	❐
rfit2_6	Das Orchester übertönte den Chor an vielen Stellen.	❐	❐	❐	❐	❐	❐	❐
vfit2_8	Die Balance zwischen dem Orchester und den Sängern war sehr gut gelungen, so dass weder das Orchester noch die Sänger „zugedeckt“ wurden.	❐	❐	❐	❐	❐	❐	❐
vrfit2_9	Es schien, als seien die Sänger vom Orchester regelrecht „überflutet“ worden.	❐	❐	❐	❐	❐	❐	❐

Kennziffer des Items	***Bitte beurteilen Sie die Stimmigkeit der szenischen Aspekte.***	Trifft voll und ganz zu	Trifft eher zu	Teils, teils	Trifft eher nicht zu	Trifft über-haupt nicht zu	Habe ich nicht beach-tet	Kann ich nicht beur-teilen
fit3A1	Das Bühnenbild unterstützte die Aufführung in Bezug auf die Atmosphäre.	❐	❐	❐	❐	❐	❐	❐
vfit2S_1	Das Bühnenbild unterstützte die Aufführung in Bezug auf die Handlung und die Darsteller.	❐	❐	❐	❐	❐	❐	❐
fit2S_2	Dekoration (Kulissen, Möbel) und Requisiten unterstützten die Handlung der zentralen Darsteller.	❐	❐	❐	❐	❐	❐	❐
vfit3A2	Die Ausstattung der Aufführung (Bühnenbild, Kostüme, Dekoration, Requisiten, Licht) war insgesamt ästhetisch sehr gelungen.	❐	❐	❐	❐	❐	❐	❐
fit3A4	Kostüme und Bühnenbild waren so aufeinander abgestimmt, dass eine „stimmige“ Atmosphäre entstand.	❐	❐	❐	❐	❐	❐	❐
rfit3A3	Die Beleuchtung wirkte häufig unpassend.	❐	❐	❐	❐	❐	❐	❐
fit3A5	Bühnenbild, Requisiten und Licht erzeugten insgesamt eine „stimmige“ Atmosphäre.	❐	❐	❐	❐	❐	❐	❐
vrfit2S_4	Die Ausstattung der Aufführung (Kostüme, Requisiten und Bühnenbild) ließ die szenische Leistung der Sänger (Solisten und Chor) wie einen Fremdkörper erscheinen.	❐	❐	❐	❐	❐	❐	❐

Kennziffer des Items	*(Fortsetzung)* *Bitte beurteilen Sie die Stimmigkeit der szenischen Aspekte.*	Trifft voll und ganz zu	Trifft eher zu	Teils, teils	Trifft eher nicht zu	Trifft über-haupt nicht zu	Habe ich nicht beach-tet	Kann ich nicht beur-teilen
vfit2S_3	Die schauspielerische Leistung der Sänger und die Ausstattung der Aufführung (Kostüme, Requisiten und Bühnenbild) ergaben insgesamt ein „stimmiges“ Bild.	❐	❐	❐	❐	❐	❐	❐

Kennziffer des Items	*Bitte beurteilen Sie die Stimmigkeit zwischen Musik und Szene.*	Trifft voll und ganz zu	Trifft eher zu	Teils, teils	Trifft eher nicht zu	Trifft über-haupt nicht zu	Habe ich nicht beach-tet	Kann ich nicht beur-teilen
vfit1_1	Die „Musik“ (Orchester, Chor und Solisten) war in dieser Aufführung insgesamt gut auf die „Szene“ (Bühnenbild, Dekoration, Requisiten, Kostüme und Beleuchtung) abgestimmt.	❐	❐	❐	❐	❐	❐	❐
rfit1_2	Die szenische Gestaltung wirkte häufig ohne Bezug zum musikalischen Geschehen.	❐	❐	❐	❐	❐	❐	❐
fit1_3	Die Inszenierung (Spiel der Darsteller, Bühnenbild, Dekoration, Requisiten, Kostüme, Beleuchtung) wurde dem Charakter der Musik gerecht.	❐	❐	❐	❐	❐	❐	❐
fit1_6	Die darstellerischen Leistungen der Solisten haben die musikalische Atmosphäre der Aufführung eindrucksvoll unterstrichen.	❐	❐	❐	❐	❐	❐	❐
fit1_7	Musikalische Gestaltung und szenisches Geschehen waren in dieser Aufführung insgesamt wie „aus einem Guss“.	❐	❐	❐	❐	❐	❐	❐
vfit1_8	Die szenische Ausstattung des Stückes (Bühnenbild, Kostüme, Dekoration) hat die musikalische Atmosphäre der Aufführung sehr gut unterstützt.	❐	❐	❐	❐	❐	❐	❐

D. abschließende Fragen

Kennziffer des Items	*Wie haben Sie die folgenden Aspekte heute Abend wahrgenommen?*	Trifft voll und ganz zu	Trifft eher zu	Teils, teils	Trifft eher nicht zu	Trifft über-haupt nicht zu
mRR1	Der Saal, in dem die Aufführung statt fand, hat eine sehr gute Akustik.	❐	❐	❐	❐	❐
mRR2	Mein Sitz war sehr bequem.	❐	❐	❐	❐	❐
mRR3	Mein Sitz bot eine sehr gute Sicht auf die Bühne.	❐	❐	❐	❐	❐
mRR4	Während der Aufführung kam es zu störenden Ablenkungen (Geräusche, etc).	❐	❐	❐	❐	❐
mRR5	Die Temperatur im Zuschauerraum war sehr angenehm.	❐	❐	❐	❐	❐
mRR6	Mit dem Service des Hauses (Bewirtung, Garderobe, Parkplatz, etc) war ich sehr zufrieden.	❐	❐	❐	❐	❐
mRR7	Die Räumlichkeiten, in denen sich das Publikum vor und nach der Vorstellung bzw. während der Pause aufhält, sind ansprechend gestaltet und laden zum Verweilen ein.	❐	❐	❐	❐	❐
mRR8	Die Atmosphäre im Theater ist sehr angenehm.	❐	❐	❐	❐	❐
mRR9	*Bitte beantworten Sie diese Frage nur, wenn Sie den Fragebogen nach der Aufführung ausfüllen.* Für meinen Geschmack war die Aufführung deutlich zu lang.	❐	❐	❐	❐	❐

(Fortsetzung) D. Abschließende Fragen

RS1: Wie alt sind Sie? _____

RS2: Ihr Geschlecht? ❒ männlich ❒ weiblich

mRS3: Welches ist Ihr höchster Schulabschluss?

❒ ohne Abschluss ❒ Abitur ❒ Universitätsabschluss
❒ Volks-/Hauptschule ❒ Ausbildung
❒ Mittlere Reife ❒ Fachhochschulabschluss

RS4: Wie groß ist Ihre Erfahrung in Bezug auf die Oper?

❒ Ich bin absoluter Laie.
❒ Ich bin Laie, habe aber etwas Opernerfahrung.
❒ Die Oper gehört zu meinen Hobbies; ich gehe häufiger in Aufführungen.
❒ Ich würde mich schon als „Halbprofi“ bezeichnen.
❒ Ich bin Experte/-in.

mRS5: Wie oft pro Jahr besuchen Sie eine Opernaufführung (im Durchschnitt)?

_____ Mal

mRS6_1: Haben Sie ein Abonnement des Anhaltischen Theaters Dessau?

❒ ja ❒ nein

mRS6_2: Wenn Sie mit ‚ja’ geantwortet haben: Seit wann haben Sie Ihr Abonnement?

seit _____ Spielzeiten

mRS7: Beherrschen Sie ein Instrument?

❒ sehr gut ❒ ziemlich gut ❒ ein bisschen ❒ kaum ❒ überhaupt nicht

RS8: Wie gut haben Sie das aufgeführte Stück bereits zuvor gekannt?

❒ sehr gut ❒ ziemlich gut ❒ ein bisschen ❒ kaum ❒ überhaupt nicht

mEP3_1: Wie war Ihre Stimmung <u>vor</u> der Aufführung?

sehr schlecht gelaunt ❒ ❒ ❒ ❒ ❒ ❒ ❒ super gelaunt

mEP3_2: Wie war Ihre Stimmung <u>während</u> der Aufführung?

sehr schlecht gelaunt ❒ ❒ ❒ ❒ ❒ ❒ ❒ super gelaunt

mEP3_3: Wie war Ihre Stimmung nach der Aufführung?

sehr schlecht gelaunt ❒ ❒ ❒ ❒ ❒ ❒ ❒ super gelaunt

Welche Akteure sind Ihrer Meinung nach am wichtigsten für die künstlerische Qualität einer Opernaufführung (allgemein, d.h. unabhängig von dieser Aufführung)?

Kennziffer des Items		Sehr wichtig	Ziemlich wichtig	Teils, teils	Eher unwichtig	Sehr unwichtig
mW1	Orchester	❒	❒	❒	❒	❒
mW2	Solisten	❒	❒	❒	❒	❒
mW3	Chor	❒	❒	❒	❒	❒
mW4	Dirigent	❒	❒	❒	❒	❒
mW5	Regisseur	❒	❒	❒	❒	❒
mW6	Bühnenbildner	❒	❒	❒	❒	❒
mW7	Techniker	❒	❒	❒	❒	❒
mW8	Beleuchter	❒	❒	❒	❒	❒

E. Allerletzte Frage

RS9: Ich habe den Fragebogen ausgefüllt …

❒ in der Pause. ❒ nach der Aufführung (im Theater). ❒ zu Hause.

Bitte geben Sie den Fragebogen in einen der Kästen am Ausgang oder senden Sie den Fragebogen in einem Freiumschlag an uns zurück.

Vielen Dank für Ihre Teilnahme an der Untersuchung

Anhang B: Fragebogen zur Bewertung von Aufführungen im Musiktheater – Renz (2006)

Sabine Renz
Lehrstuhl für Management, insbesondere Strategie und Führung
Fachbereich Politik- und Verwaltungswissenschaft
Universität Konstanz

A. Gesamteindruck

*In diesem Teil des Fragebogens interessiert uns, wie Dir die Aufführung **insgesamt** gefallen hat.*

1. Gesamturteil	Trifft voll und ganz zu	Trifft eher zu	Teils, teils	Trifft eher nicht zu	Trifft über-haupt nicht zu	Kann ich nicht beantworten
Die Aufführung hat mir insgesamt gut gefallen.	❐	❐	❐	❐	❐	❐
Die Aufführung war in meinen Augen ein ziemlicher Flop.	❐	❐	❐	❐	❐	❐
Die Produktion hatte hohes künstlerisches und technisches Niveau.	❐	❐	❐	❐	❐	❐
Ich würde erwarten, dass die Produktion von der Kritik gut aufgenommen und besprochen wurde.	❐	❐	❐	❐	❐	❐
Der Besuch der Aufführung hat mir nicht viel gebracht.	❐	❐	❐	❐	❐	❐
Die Aufführung war für mich eine persönliche Bereicherung.	❐	❐	❐	❐	❐	❐
Die Aufführung war insgesamt anregend und nie langweilig.	❐	❐	❐	❐	❐	❐
Es wäre kein Verlust, wenn diese Produktion schon bald aus dem Spielplan genommen würde.	❐	❐	❐	❐	❐	❐
Die Produktion hat gezeigt, welch eine hohe Kunstform die Oper sein kann.	❐	❐	❐	❐	❐	❐

	Trifft voll und ganz zu	Trifft eher zu	Teils, teils	Trifft eher nicht zu	Trifft über-haupt nicht zu	Kann ich nicht beantworten
Die Darbietung hat mich gefühlsmäßig stark berührt.	❐	❐	❐	❐	❐	❐
Die Aufführung hat mich insgesamt ziemlich kalt gelassen.	❐	❐	❐	❐	❐	❐
Ich war so begeistert, dass ich mir die Produktion demnächst gerne noch einmal ansehen würde.	❐	❐	❐	❐	❐	❐
Ich würde die Produktion auch Freunden und Bekannten empfehlen.	❐	❐	❐	❐	❐	❐
Ich würde die Produktion nicht unbedingt weiterempfehlen.	❐	❐	❐	❐	❐	❐
Ich würde mir die Produktion auch mit einer Freikarte nicht noch einmal ansehen wollen.	❐	❐	❐	❐	❐	❐
Diese Aufführung wird mich bestimmt noch länger beschäftigen.	❐	❐	❐	❐	❐	❐

2. Die Inszenierung (...)	Trifft voll und ganz zu	Trifft eher zu	Teils, teils	Trifft eher nicht zu	Trifft über-haupt nicht zu	Habe ich nicht beach-tet	Kann ich nicht beur-teilen
Das Gesamtkonzept der Inszenierung war für mich gut erkennbar.	❐	❐	❐	❐	❐	❐	❐
(…) war so, dass der Leitgedanke des Regiekonzepts verständlich wurde.	❐	❐	❐	❐	❐	❐	❐
(...) ergab in meinen Augen keine schlüssige Botschaft.	❐	❐	❐	❐	❐	❐	❐
Das Regiekonzept dieser Inszenierung hat mir überhaupt nicht gefallen.	❐	❐	❐	❐	❐	❐	❐
Der Grundgedanke dieser Inszenierung hat mich absolut überzeugt.	❐	❐	❐	❐	❐	❐	❐
(…) wird nach meinem Eindruck dem Stück bzw. der Intention des Komponisten voll gerecht.	❐	❐	❐	❐	❐	❐	❐
(...) war zwar in sich schlüssig, gefiel mir aber trotzdem nicht.	❐	❐	❐	❐	❐	❐	❐

	Trifft voll und ganz zu	**Trifft eher zu**	**Teils, teils**	**Trifft eher nicht zu**	**Trifft über-haupt nicht zu**		**Habe ich nicht beach-tet**	**Kann ich nicht beur-teilen**
Bühnenbild und Dekoration lenkten durch eine Fülle von Einzelheiten von der Handlung ab.	❐	❐	❐	❐	❐		❐	❐
Auf der Bühne war oft soviel los, dass man nicht wusste, wo man eigentlich hinschauen sollte.	❐	❐	❐	❐	❐		❐	❐
Es war oft schwierig, den genauen Fortgang der Handlung zu verfolgen.	❐	❐	❐	❐	❐		❐	❐
Die zu häufige Bewegung der Darsteller im Raum lenkte von der eigentlichen Handlung ab.	❐	❐	❐	❐	❐		❐	❐
Bühnenbild und Dekoration (Kulissen, Möbel) wurden zu sparsam eingesetzt.	❐	❐	❐	❐	❐		❐	❐
Die Solisten standen häufig unmotiviert auf der Bühne herum.	❐	❐	❐	❐	❐		❐	❐
Bühnenbild, Kostüme und Licht sorgten insgesamt für eindringliche Bilder.	❐	❐	❐	❐	❐		❐	❐
Ich konnte während der Aufführung alles um mich herum vergessen und mich auf das Bühnengeschehen gut einlassen.	❐	❐	❐	❐	❐		❐	❐
Die Aktion der Personen auf der Bühne passte ideal zur Handlung des Stücks.	❐	❐	❐	❐	❐		❐	❐
Die Aufführung war eine wohltuende Abwechslung von der „Opernroutine“.	❐	❐	❐	❐	❐		❐	❐
Die Inszenierung arbeitete mit vielen ungewohnten Stilmitteln.	❐	❐	❐	❐	❐		❐	❐
Die Inszenierung war mir insgesamt zu neuartig und modern.	❐	❐	❐	❐	❐		❐	❐
Die Art der Aufführung hat mich ziemlich irritiert.	❐	❐	❐	❐	❐		❐	❐
Die Aufführung hat mir insgesamt kaum etwas Neues geboten.	❐	❐	❐	❐	❐		❐	❐
Die Aufführung war äußerst erfrischend.	❐	❐	❐	❐	❐		❐	❐
Ich hätte mir mehr Originalität gewünscht.	❐	❐	❐	❐	❐		❐	❐

B. Bewertung der Einzelaspekte

In diesem Teil des Fragebogens interessiert uns, wie Du das Orchester, die Solisten, den Chor und den Dirigenten der Opernaufführung jeweils für sich genommen bewertest.

I. Musikalische Leistung des Ensembles

1. Orchester	**Trifft voll und ganz zu**	**Trifft eher zu**	**Teils, teils**	**Trifft eher nicht zu**	**Trifft über-haupt nicht zu**	**Habe ich nicht beach-tet**	**Kann ich nicht beur-teilen**
Unterschiede in der Lautstärke wurden vom Orchester gut herausgearbeitet.	❐	❐	❐	❐	❐	❐	❐
Dem Orchester ist es gelungen, den ganzen Abend über die musikalische Spannung zu halten.	❐	❐	❐	❐	❐	❐	❐
Bei Änderungen des Tempos reagierte das Orchester flexibel und souverän.	❐	❐	❐	❐	❐	❐	❐
Das Orchester wirkte insgesamt häufig überfordert.	❐	❐	❐	❐	❐	❐	❐
Es gelang dem Orchester, die unterschiedlichen Tempi überzeugend herauszuarbeiten.	❐	❐	❐	❐	❐	❐	❐
Das Orchester spielte mit großer Spannung und Intensität.	❐	❐	❐	❐	❐	❐	❐
Das Orchester spielte sehr lebendig.	❐	❐	❐	❐	❐	❐	❐
Der Orchesterklang überzeugte durch hohe Klarheit und Durchsichtigkeit.	❐	❐	❐	❐	❐	❐	❐
Das Orchester spielte sehr präzise.	❐	❐	❐	❐	❐	❐	❐

	Trifft voll und ganz zu	**Trifft eher zu**	**Teils, teils**	**Trifft eher nicht zu**	**Trifft über-haupt nicht zu**		**Habe ich nicht beach-tet**	**Kann ich nicht beur-teilen**
Der Solist/die Solistin sang bei seinen/ihren Soli mit einem hohen Nuancenreichtum.	❐	❐	❐	❐	❐		❐	❐
Die Stimme des Solisten/der Solistin besaß wenig klangliche Schönheit.	❐	❐	❐	❐	❐		❐	❐
Der Solist/die Solistin sprach seinen/ihren Text sehr deutlich aus.	❐	❐	❐	❐	❐		❐	❐
Die Stimme des Solisten/der Solistin war ideal für die Anforderungen der Partie.	❐	❐	❐	❐	❐		❐	❐
Der Solist/die Solistin brachte die Unterschiede in der Lautstärke seiner/ihrer Partie gut zur Geltung.	❐	❐	❐	❐	❐		❐	❐
Der Solist/die Solistin war seiner/ihrer Partie musikalisch insgesamt gut gewachsen.	❐	❐	❐	❐	❐		❐	❐
Stimmlich vermittelte der Solist/die Solistin die emotionale Atmosphäre der von ihm/ihr gesungenen Partie gut.	❐	❐	❐	❐	❐		❐	❐
Der Solist/die Solistin wirkte stimmlich häufig angestrengt bzw. krampfhaft bemüht.	❐	❐	❐	❐	❐		❐	❐
Der Solist/die Solistin wurde dem Anspruch der Rolle musikalisch gerecht.	❐	❐	❐	❐	❐		❐	❐

3. Chor	**Trifft voll und ganz zu**	**Trifft eher zu**	**Teils, teils**	**Trifft eher nicht zu**	**Trifft über-haupt nicht zu**		**Habe ich nicht beach-tet**	**Kann ich nicht beur-teilen**
Der Chor brachte Unterschiede in der Lautstärke sehr gut zur Geltung.	❐	❐	❐	❐	❐		❐	❐
Bei den Chorpartien mangelte es häufig an rhythmischer Präzision.	❐	❐	❐	❐	❐		❐	❐
Der Chor überzeugte durch klangliche Brillanz.	❐	❐	❐	❐	❐		❐	❐

	Trifft voll und ganz zu	Trifft eher zu	Teils, teils	Trifft eher nicht zu	Trifft über-haupt nicht zu	Habe ich nicht beach-tet	Kann ich nicht beur-teilen
Der Chor zeigte sich dem Stück in jeder Hinsicht gewachsen.	❐	❐	❐	❐	❐	❐	❐
Die Chorsänger sprachen ihren Text sehr deutlich aus.	❐	❐	❐	❐	❐	❐	❐

4. Dirigent	Trifft voll und ganz zu	Trifft eher zu	Teils, teils	Trifft eher nicht zu	Trifft über-haupt nicht zu	Habe ich nicht beach-tet	Kann ich nicht beur-teilen
Der Dirigent hatte das Orchester spürbar gut „im Griff“.	❐	❐	❐	❐	❐	❐	❐
Die Tempi, die der Dirigent gewählt hat, wirkten insgesamt überzeugend.	❐	❐	❐	❐	❐	❐	❐
Der Dirigent schien kein Konzept zu haben: Der Aufführung fehlte die musikalische Linie über die verschiedenen Nummern hinweg.	❐	❐	❐	❐	❐	❐	❐
Der Dirigent motivierte das Orchester zu musikalischen Höchstleistungen.	❐	❐	❐	❐	❐	❐	❐
Orchester und Dirigent schienen sich gegenseitig zu beflügeln.	❐	❐	❐	❐	❐	❐	❐
Der Dirigent verstand es, die Balance zwischen Orchestergraben und Sängern gut zu halten.	❐	❐	❐	❐	❐	❐	❐
Der Dirigent besitzt nach meinem Wissen ein hohes Renommee.	❐	❐	❐	❐	❐	❐	❐
Der Dirigent verlangte den Solisten geradezu unmenschliche Tempi ab.	❐	❐	❐	❐	❐	❐	❐
Der Dirigent inspirierte das gesamte Ensemble zu musikalischen Höchstleistungen.	❐	❐	❐	❐	❐	❐	❐
Der Dirigent zeigte sich dem Werk in jeder Hinsicht gewachsen.	❐	❐	❐	❐	❐	❐	❐

II. Szenische/darstellerische Leistung

1. Solist/-in *Bitte beziehe die folgenden Fragen wieder ausschließlich auf die eine männliche oder weibliche Hauptrolle, welche Dir zu Beginn gezeigt wurde.*	**Trifft voll und ganz zu**	**Trifft eher zu**	**Teils, teils**	**Trifft eher nicht zu**	**Trifft über-haupt nicht zu**	**Habe ich nicht beach-tet**	**Kann ich nicht beur-teilen**
Der Solist/die Solistin spielte die Figur eindringlich heraus.	❐	❐	❐	❐	❐	❐	❐
Die körperliche Erscheinung des Solisten/der Solistin passte sehr gut zu seiner/ihrer gespielten Rolle.	❐	❐	❐	❐	❐	❐	❐
Der Solist/die Solistin besitzt nach meinem Wissen ein hohes Renommee.	❐	❐	❐	❐	❐	❐	❐
Der Solist/die Solistin lebte die Rolle so, dass ich mich gut mit ihr identifizieren konnte.	❐	❐	❐	❐	❐	❐	❐
Der Solist/die Solistin verkörperte die Charaktereigenschaften der Figur sehr überzeugend.	❐	❐	❐	❐	❐	❐	❐
Das äußere Erscheinungsbild (Kostüm und Maske) des solistischen Darstellers/der solistischen Darstellerin passte gut zum Inhalt der Rolle.	❐	❐	❐	❐	❐	❐	❐
Der Darsteller/die Darstellerin war äußerlich sehr attraktiv.	❐	❐	❐	❐	❐	❐	❐
Mimik, Gestik und Bewegungen des solistischen Darstellers/der solistischen Darstellerin passten oft nicht zum szenischen Geschehen.	❐	❐	❐	❐	❐	❐	❐
Die Führung der Figuren (Verteilung und Bewegung der Solisten im Raum) gefiel mir insgesamt gut.							

2. Chor	**Trifft voll und ganz zu**	**Trifft eher zu**	**Teils, teils**	**Trifft eher nicht zu**	**Trifft über-haupt nicht zu**	**Habe ich nicht beach-tet**	**Kann ich nicht beur-teilen**
Das äußere Erscheinungsbild der Chorsänger (Kostüm und Maske) passte gut zur Rolle/Partie des Chores in dieser Aufführung.	❐	❐	❐	❐	❐	❐	❐
Man konnte dem Chor durch sein Auftreten und Aussehen seine Rolle voll abnehmen.	❐	❐	❐	❐	❐	❐	❐
Das äußere Erscheinungsbild der Chorsänger (Kostüm und Maske) war ästhetisch befriedigend.	❐	❐	❐	❐	❐	❐	❐
Die Anordnung und Bewegung des Chores im Raum passte gut zur Rolle des Chores.	❐	❐	❐	❐	❐	❐	❐
Der Chor war mir optisch zu sehr im Hintergrund.	❐	❐	❐	❐	❐	❐	❐
Schauspielerisch fand ich den Chor insgesamt sehr überzeugend.							

3. Ausstattung *(Bühne, Kostüme, Dekoration und Requisiten)*	**Trifft voll und ganz zu**	**Trifft eher zu**	**Teils, teils**	**Trifft eher nicht zu**	**Trifft über-haupt nicht zu**	**Habe ich nicht beach-tet**	**Kann ich nicht beur-teilen**
Das Bühnenbild hat mir insgesamt gut gefallen.	❐	❐	❐	❐	❐	❐	❐
Der Einsatz der Bühnentechnik hat mit gut gefallen.	❐	❐	❐	❐	❐	❐	❐
Die Kostüme haben mir insgesamt gut gefallen.	❐	❐	❐	❐	❐	❐	❐

C. Bewertung des Zusammenspiels

*In diesem Teil des Fragebogens geht es uns nicht darum, wie Dir die einzelnen Teilaspekte für sich gefallen haben, sondern darum, wie Du das **„Zusammenspiel“** dieser Aspekte empfunden hast.*

1. Zusammenspiel innerhalb des Orchesters	**Trifft voll und ganz zu**	**Trifft eher zu**	**Teils, teils**	**Trifft eher nicht zu**	**Trifft über-haupt nicht zu**	**Habe ich nicht beach-tet**	**Kann ich nicht beur-teilen**
Die verschiedenen Instrumentengruppen des Orchesters (Streicher, Bläser, Schlagwerk) waren in Bezug auf **die Lautstärke** gut aufeinander abgestimmt.	❐	❐	❐	❐	❐	❐	❐
Die verschiedenen Instrumentengruppen des Orchesters (Streicher, Bläser, Schlagwerk) waren in Bezug auf **das Tempo** gut aufeinander abgestimmt.	❐	❐	❐	❐	❐	❐	❐
Die verschiedenen Instrumentengruppen des Orchesters (Streicher, Bläser, Schlagwerk) waren in Bezug auf **die Klangfarbe** gut aufeinander abgestimmt.	❐	❐	❐	❐	❐	❐	❐
Es kam oft vor, dass eine Instrumentengruppe in störender Weise das ganze Orchester dominierte.	❐	❐	❐	❐	❐	❐	❐
Das Orchester spielte „wie aus einem Guss“.	❐	❐	❐	❐	❐	❐	❐
Das Blech klang häufig zu laut und zu hart.	❐	❐	❐	❐	❐	❐	❐
Die Streichergruppe war in sich gut synchronisiert.	❐	❐	❐	❐	❐	❐	❐
Die Orchestermusiker waren in Bezug auf die Einsätze gut aufeinander abgestimmt.	❐	❐	❐	❐	❐	❐	❐
Die solistischen Instrumente im Orchester (z.B. Holzbläser) klangen sehr ausdrucksvoll.	❐	❐	❐	❐	❐	❐	❐

2. Zusammenspiel innerhalb des Chores	**Trifft voll und ganz zu**	**Trifft eher zu**	**Teils, teils**	**Trifft eher nicht zu**	**Trifft über-haupt nicht zu**		**Habe ich nicht beach-tet**	**Kann ich nicht beur-teilen**
Der Chor wirkte in sich klanglich sehr homogen.	❐	❐	❐	❐	❐		❐	❐
Die Chorstimmen waren in Bezug auf das Tempo gut aufeinander abgestimmt.	❐	❐	❐	❐	❐		❐	❐
Der Chorklang wirkte klanglich ausgewogen, so dass nie einzelne Stimmen einzelner Personen deutlich herauszuhören waren.	❐	❐	❐	❐	❐		❐	❐

3. Zusammenspiel der Solisten (Hauptdarsteller) *Bitte beziehe in diesem Teil, entgegen dem bisherigen Prinzip, Deine Aussagen auf das „Zusammenspiel“ **aller** Solistinnen und Solisten in den Hauptrollen **untereinander.***	**Trifft voll und ganz zu**	**Trifft eher zu**	**Teils, teils**	**Trifft eher nicht zu**	**Trifft über-haupt nicht zu**		**Habe ich nicht beach-tet**	**Kann ich nicht beur-teilen**
Die Solisten waren klanglich untereinander gut abgestimmt.	❐	❐	❐	❐	❐		❐	❐
Die Solisten waren in Bezug auf das Tempo untereinander gut abgestimmt.	❐	❐	❐	❐	❐		❐	❐
Es kam häufig vor, dass die Solisten mehr „gegeneinander“ als miteinander musizierten.	❐	❐	❐	❐	❐		❐	❐
Die Solisten passten schauspielerisch von ihrer Ausstrahlung her gut zueinander.	❐	❐	❐	❐	❐		❐	❐
Das Solistenensemble wirkte musikalisch untereinander ausgesprochen harmonisch.	❐	❐	❐	❐	❐		❐	❐
Man hatte den Eindruck, dass sich die Solisten gegenseitig „an die Wand spielen/singen.“	❐	❐	❐	❐	❐		❐	❐

4. Zusammenspiel von Orchester, Solisten und Chor	**Trifft voll und ganz zu**	**Trifft eher zu**	**Teils, teils**	**Trifft eher nicht zu**	**Trifft über-haupt nicht zu**		**Habe ich nicht beach-tet**	**Kann ich nicht beur-teilen**
Die Solisten hatten genügend Strahlkraft und Volumen, um gegen das Orchester anzukommen.	❐	❐	❐	❐	❐		❐	❐
Orchester, Chor und Solisten waren in Bezug auf die Lautstärke gut aufeinander abgestimmt.	❐	❐	❐	❐	❐		❐	❐
Zwischen Orchester, Chor und Solisten gab es rhythmische Unstimmigkeiten.	❐	❐	❐	❐	❐		❐	❐
Orchester und Chor passten vom Tempo her gut zusammen: Das Orchester war weder langsamer noch schneller als der Chor.	❐	❐	❐	❐	❐		❐	❐
Das Orchester übertönte den Chor an vielen Stellen.	❐	❐	❐	❐	❐		❐	❐
Orchester, Chor und Solisten verschmolzen musikalisch zu einem harmonischen Klangbild.	❐	❐	❐	❐	❐		❐	❐
Die Balance zwischen dem Orchester und den Sängern war gelungen, so dass weder das Orchester noch die Sänger „zugedeckt“ wurden.	❐	❐	❐	❐	❐		❐	❐
Es schien, als seien die Sänger vom Orchester „überflutet“ worden.	❐	❐	❐	❐	❐		❐	❐

5. Stimmigkeit der szenischen Aspekte	**Trifft voll und ganz zu**	**Trifft eher zu**	**Teils, teils**	**Trifft eher nicht zu**	**Trifft über-haupt nicht zu**		**Habe ich nicht beach-tet**	**Kann ich nicht beur-teilen**
Das Bühnenbild unterstützte die Aufführung in Bezug auf die Atmosphäre.	❐	❐	❐	❐	❐		❐	❐
Das Bühnenbild unterstützte die Aufführung in Bezug auf die Handlung und die Figuren/ Darsteller.	❐	❐	❐	❐	❐		❐	❐
Dekoration (Kulissen, Möbel) und Requisiten unterstützten die Handlung der zentralen Darsteller.	❐	❐	❐	❐	❐		❐	❐
Die Ausstattung der Aufführung (Bühnenbild, Kostüme, Dekoration, Requisiten, Licht) war insgesamt ästhetisch gelungen.	❐	❐	❐	❐	❐		❐	❐
Die Beleuchtung wirkte häufig unpassend.	❐	❐	❐	❐	❐		❐	❐

	Trifft voll und ganz zu	Trifft eher zu	Teils, teils	Trifft eher nicht zu	Trifft über-haupt nicht zu		Habe ich nicht beach-tet	Kann ich nicht beur-teilen
Kostüme und Bühnenbild waren so aufeinander abgestimmt, dass eine „stimmige“ Atmosphäre entstand.	❐	❐	❐	❐	❐		❐	❐
Bühnenbild, Requisiten und Licht erzeugten insgesamt eine „stimmige“ Atmosphäre.	❐	❐	❐	❐	❐		❐	❐
Die schauspielerische Leistung der Sänger und die Ausstattung der Oper (Kostüme, Requisiten und Bühnenbild) ergaben insgesamt ein stimmiges Bild.	❐	❐	❐	❐	❐		❐	❐
Die Ausstattung der Oper (Kostüme, Requisiten und Bühnenbild) ließ die szenische Leistung der Sänger (Solisten und Chor) wie einen Fremdkörper erscheinen.	❐	❐	❐	❐	❐		❐	❐

6. Stimmigkeit zwischen Musik und Szene	Trifft voll und ganz zu	Trifft eher zu	Teils, teils	Trifft eher nicht zu	Trifft über-haupt nicht zu		Habe ich nicht beach-tet	Kann ich nicht beur-teilen
Die „Musik“ (Orchester, Chor und Solisten) war in dieser Aufführung insgesamt gut auf die „Szene“ (Bühnenbild, Dekoration, Requisiten, Kostüme, Beleuchtung) abgestimmt.	❐	❐	❐	❐	❐		❐	❐
Die szenische Gestaltung wirkte häufig ohne Bezug zum musikalischen Geschehen.	❐	❐	❐	❐	❐		❐	❐
Die Inszenierung (Spiel der Darsteller, Bühnenbild, Dekoration, Requisiten, Kostüme, Beleuchtung) wurde dem Charakter der Musik gerecht.	❐	❐	❐	❐	❐		❐	❐
Die Inszenierung stand in krassem Widerspruch zu der musikalischen Interpretation des Werkes.	❐	❐	❐	❐	❐		❐	❐
Die Aufführung lief insgesamt dem Charakter des Werkes zuwider.	❐	❐	❐	❐	❐		❐	❐
Die darstellerischen Leistungen der Solisten haben die musikalische Atmosphäre der Aufführung eindrucksvoll unterstrichen.	❐	❐	❐	❐	❐		❐	❐

	Trifft voll und ganz zu	**Trifft eher zu**	**Teils, teils**	**Trifft eher nicht zu**	**Trifft über-haupt nicht zu**		**Habe ich nicht beach-tet**	**Kann ich nicht beur-teilen**
Musikalische Gestaltung und szenisches Geschehen waren in dieser Aufführung insgesamt „aus einem Guss“.	❐	❐	❐	❐	❐		❐	❐
Die szenische Ausstattung des Stückes (Bühnenbild, Kostüme, Dekoration) hat die musikalische Atmosphäre der Aufführung gut unterstützt.	❐	❐	❐	❐	❐		❐	❐

Wie war Deine Stimmung während der Vorführung?

Falls wir nicht alles gefragt haben sollten, was Dir für die Bewertung der gesehenen Aufführung wichtig ist:

Besonders **gut** hat mir an dieser Aufführung gefallen:

_____

Besonders **schlecht** fand ich an dieser Aufführung:

_____

Hast Du das Werk zuvor gekannt?

ja		**nein**	
gut	ein bisschen	kaum	überhaupt nicht
❐	❐	❐	❐

Experiment zur Bewertung von Aufführungen im Musiktheater

Persönliche Angaben:

Wie alt bist Du? _____

Geschlecht: männlich ❐ weiblich ❐

Falls Du Student/-in bist:
Welchen Studiengang studierst Du?

_____

In welchem Fachsemester bist Du?

_____

Wie würdest Du Deine Erfahrung in Bezug auf Musiktheater einschätzen?
(bitte ankreuzen)

❐ Ich bin absoluter Laie/-in.
❐ Ich bin Laie/-in, habe aber etwas Musiktheatererfahrung.
❐ Musiktheater gehört zu meinen „Hobbies"; ich gehe häufiger in Aufführungen.
❐ Ich würde mich schon als „Halbprofi" bezeichnen.
❐ Experte/-in.

Wie schätzt Du Dich selber ein? **Ich bin…**	**Trifft voll und ganz zu**	**Trifft eher zu**	**Teils, teils**	**Trifft eher nicht zu**	**Trifft über-haupt nicht zu**
... zurückhaltend	❐	❐	❐	❐	❐
… eher ruhig	❐	❐	❐	❐	❐
… kommunikativ, gesprächig	❐	❐	❐	❐	❐
… einfallsreich	❐	❐	❐	❐	❐
… eher unorganisiert	❐	❐	❐	❐	❐

	Trifft voll und ganz zu	**Trifft eher zu**	**Teils, teils**	**Trifft eher nicht zu**	**Trifft über-haupt nicht zu**
… eher faul	❐	❐	❐	❐	❐
… jemand, der aus sich herausgeht, gesellig ist	❐	❐	❐	❐	❐
… jemand, der sich manchmal gehemmt fühlt, schüchtern ist	❐	❐	❐	❐	❐
… jemand, der gern reflektiert, mit Ideen spielt	❐	❐	❐	❐	❐
… jemand, der eine lebhafte Phantasie, Vorstellung hat	❐	❐	❐	❐	❐
…jemand, der gern künstlerische, ästhetische Erfahrungen schätzt	❐	❐	❐	❐	❐
… jemand, der originell ist, neue Ideen einbringt	❐	❐	❐	❐	❐
… jemand, der gründlich arbeitet	❐	❐	❐	❐	❐
… jemand, der Aufgaben wirksam und effizient erledigt	❐	❐	❐	❐	❐
… jemand, der bis zum Ende einer Aufgabe durchhält	❐	❐	❐	❐	❐
… jemand, der sich oft Sorgen macht	❐	❐	❐	❐	❐
… jemand, der entspannt ist, mit Stress gut umgehen kann	❐	❐	❐	❐	❐
… jemand, der angespannt sein kann	❐	❐	❐	❐	❐
… jemand, der leicht nervös wird	❐	❐	❐	❐	❐
… jemand, der nicht leicht aus der Ruhe zu bringen ist	❐	❐	❐	❐	❐
… jemand, der manchmal etwas grob zu anderen ist	❐	❐	❐	❐	❐
… jemand, der kalt und distanziert ist	❐	❐	❐	❐	❐
… jemand, der Streit anfängt	❐	❐	❐	❐	❐
… jemand, der rücksichtsvoll und freundlich mit anderen umgeht	❐	❐	❐	❐	❐
… jemand, der verzeihen kann	❐	❐	❐	❐	❐

Vielen Dank für Deine Teilnahme am Experiment!

Anhang C: Einfaktorielle Varianzanalysen

Tabelle 28: Einfaktorielle Varianzanalysen (Faktorvariable: Zeitpunkt des Ausfüllens)

		Quadrat-summe	Freiheits-grade	Mittel der Quadrate	F	Signifi-kanz
kQ	zwischen den Gruppen	6.993	2	3.497	3.626	.030
	innerhalb der Gruppen	108.957	113	.964		
	gesamt	115.950	115			
fit1	zwischen den Gruppen	2.856	2	1.428	2.013	.139
	innerhalb der Gruppen	70.239	99	.709		
	gesamt	73.095	101			
fit2_Musik	zwischen den Gruppen	1.398	2	.699	1.519	.224
	innerhalb der Gruppen	48.791	106	.460		
	gesamt	50.189	108			
fit2_Szene	zwischen den Gruppen	2.041	2	1.020	1.160	.318
	innerhalb der Gruppen	91.525	104	.880		
	gesamt	93.566	106			
Orchester	zwischen den Gruppen	.135	2	.067	.118	.889
	innerhalb der Gruppen	48.025	84	.572		
	gesamt	48.160	86			
MSolist	zwischen den Gruppen	.496	2	.248	.576	.564
	innerhalb der Gruppen	43.861	102	.430		
	gesamt	44.357	104			
MChor	zwischen den Gruppen	1.616	2	.808	1.342	.266
	innerhalb der Gruppen	60.824	101	.602		
	gesamt	62.440	103			

(Fortsetzung)		Quadrat-summe	Freiheits-grade	Mittel der Quadrate	F	Signifi-kanz
SSolist	zwischen den Gruppen	5.471	2	2.735	3.413	.036
	innerhalb der Gruppen	88.156	110	.801		
	gesamt	93.627	112			
SChor	zwischen den Gruppen	.891	2	.446	.361	.698
	innerhalb der Gruppen	129.614	105	1.234		
	gesamt	130.505	107			
SAustattung	zwischen den Gruppen	3.596	2	1.798	1.641	.198
	innerhalb der Gruppen	121.591	111	1.095		
	gesamt	125.187	113			
fit3_Chor	zwischen den Gruppen	.019	2	.009	.023	.978
	innerhalb der Gruppen	41.323	99	.417		
	gesamt	41.342	101			
fit3_Solisten	zwischen den Gruppen	3.184	2	1.592	3.029	.052
	innerhalb der Gruppen	56.768	108	.526		
	gesamt	59.952	110			

Anmerkungen: kQ = künstlerische Qualität, fit1 = Stimmigkeit zwischen der musikalischen und szenischen Dimension, fit2_Musik = Stimmigkeit innerhalb der musikalischen Dimension, fit2_Szene = Stimmigkeit innerhalb der szenischen Dimension, Orchester = musikalische Leistung und Zusammenspiel des Orchesters sowie Leistung des Dirigenten, MSolist = Musikalische Leistung des Solisten, MChor = Musikalische Leistung des Chores, SSolist = Szenische Leistung des Solisten, SChor = Szenischen Leistung des Chores, SAusstattung = Ausstattung (Einzelaspekte), fit3_Chor = Zusammenspiel innerhalb des Chores, fit3_Solisten = Zusammenspiel der Solisten untereinander.

Anhang D: Reliabilitäten der theoretischen Konstrukte im Vergleich

Tabelle 29: Reliabilitäten der theoretischen Konstrukte der Studien von Renz (2006), Moser (2005) und Boerner et al. (in Druck)

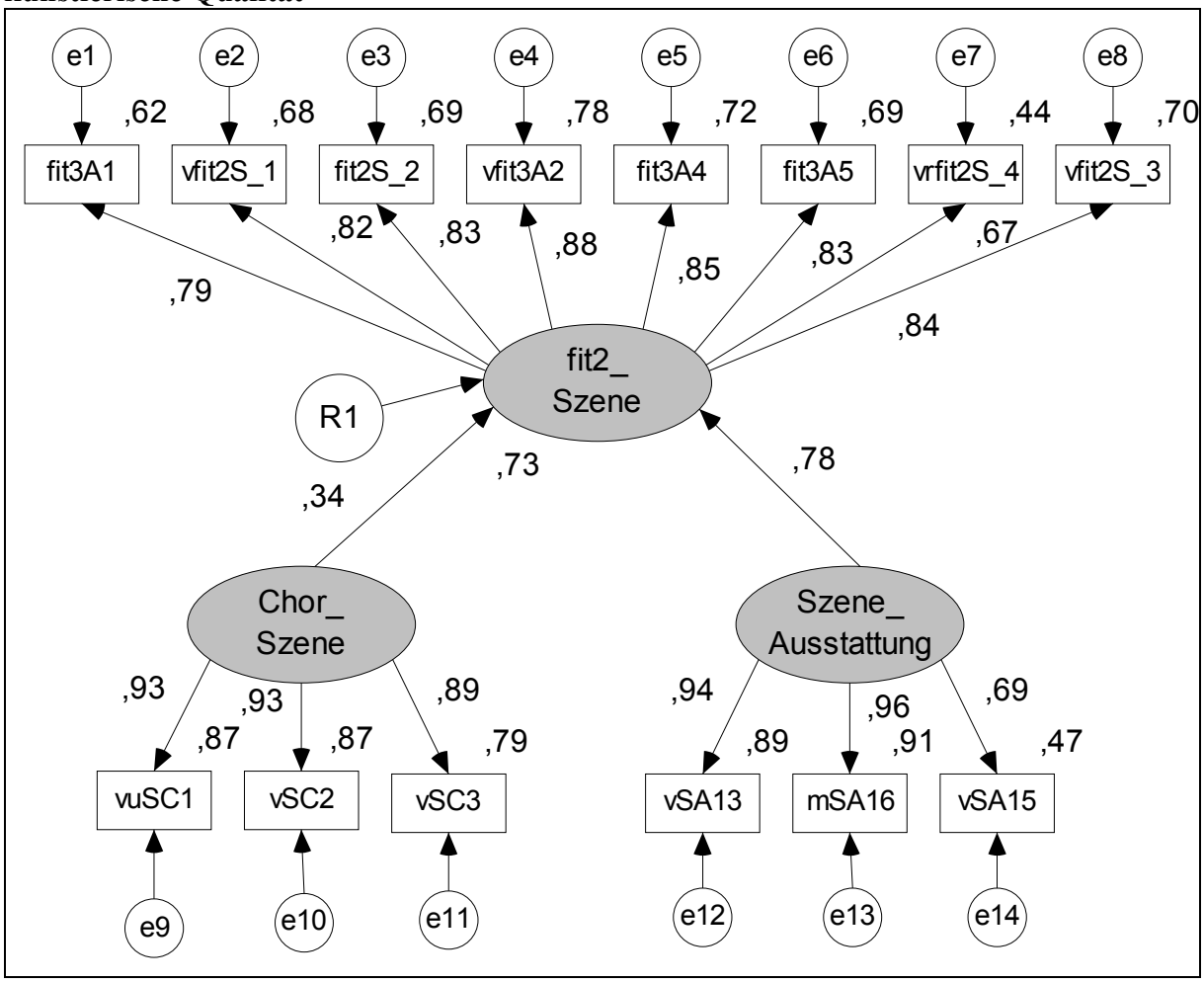
Konstrukt	Aida	La Bohème I	*La Bohème II**	Meister-singer	La forza del destino
MOrchester	.85 (3/3/9) *.82 (8/9)*	.71 (3/3/9) *.83 (8/9)*	*.83 (5/9)*	.74 (4/4/9) *.83 (5/9)*	.88 (3/3/9)
MSolisten	.65 (3/3/9) *.87 (12/13)*	.80 (5/7/9) *.81 (9/13)*	*.91 (9/13)*	.89 (8/8/9) *.85 (9/13)*	.93 (7/8/9)
MChor	-- (1/1/5) *.84 (8/8)*	.87 (4/4/5) *.89 (8/8)*	*.89 (6/8)*	.63 (3/3/5) *.91 (6/8)*	.87 (3/3/5)
MDirigent	-- (-/1/10) *.89 (8/11)*	+	+	-- (-/1/10) *.80 (14/18)*	-- (-/-/10)
SSolisten	.69 (6/7/9) *.85 (8/11)*	.81 (8/9/9) *.82 (9/11)*	*.83 (8/11)*	.88 (7/8/9) *.84 (7/11)*	.89 (8/8/9)
SChor	.85 (4/5/6) *.81 (4/5)*	.80 (4/5/6) *.79 (4/5)*	*.85 (4/5)*	.86 (5/6/7) *.90 (4/5)*	.91 (5/6/6)
SAusstattung	.85 (3/3/3)	-- (2/2/3)		-- (2/2/3)	-- (2/2/3)
fit3_Orchester	-- (-/-/9) *.83 (7/8)*	-- (-/-/9) *.88 (7/8)*	*.72 (5/8)*	-- (2/2/9) *.85 (4/8)*	-- (-/-/9)
fit3_Chor	-- (1/1/3) *.73 (4/5)*	-- (2/2/3) *.80 (5/5)*	*.71 (3/5)*	.85 (3/3/3) *.86 (3/5)*	.87 (3/3/3)
fit3_Solisten	.71 (4/5/6) *.80 (5/7)*	.79 (6/6/6) *.81 (4/7)*	*.70 (4/7)*	.73 (3/4/6) *.74 (4/7)*	.82 (4/5/6)
fit3_Ausstattung und fit2_Szene	.86 (7/8/9) *.88 (10/12)*	.91 (7/8/9) *.88 (12/12)*	*.92 (9/12)*	.92 (7/8/9) *.89 (9/12)*	.96 (7/8/9)
fit3_Ausstattung	.75 (4/4/5)	.90 (3/4/5)		.90 (3/4/5)	
fit2_Szene	.75 (3/4/4)	.79 (3/4/4)		.85 (3/4/4)	
fit2_Musik	.79 (6/6/8) *.86 (8/9)*	.84 (6/6/8) *.86 (7/9)*	*.75 (5/9)*	.83 (6/6/8) *.90 (6/9)*	.87 (6/6/8)
fit1	.78 (6/6/8) *.89 (9/9)*	.89 (6/6/8) *.84 (7/8)*	*.87 (5/8)*	.83 (6/6/8) *.88 (5/8)*	.89 (5/5/8)

(Fortsetzung) Konstrukt	Aida	La Bohème I	*La Bohème II**	Meister-singer	La forza del destino
Aufführung	*.82 (8/10)*	*.71 (9/19)*	*.81 (8/10)*	*.86 (8/10)*	
Urteil	*.89 (5/7)*	*.83 (5/7)*	*.83 (5/7)*	*.89 (5/7)*	
Gesamturteil	.92 (15/15/16)	.94 (16/16/16)		.94 (16/16/16)	.95 (16/16/16)

Anmerkungen: Untersuchung Renz (2006, Laienstichprobe, n = 39): Cronbachs Alpha (Items der Skala / Items < 30% fehlende / ursprüngliche Items); Untersuchung Moser (2005, n = 70), kursiviert: Cronbachs Alpha (Items der Skala / ursprüngliche Items); Untersuchung Boerner et al. (in Druck, n = 145): Cronbachs Alpha (Items der Skala / Items < 30% fehlende / ursprüngliche Items); fehlende Angaben ergeben sich aufgrund der Modifikation der Konstrukte bei Renz (2006); * Bei der Untersuchung Mosers (2005) wurde den Versuchspersonen zwei Inszenierungen der Oper „La Bohème“ gezeigt; + Der Dirigent war im gezeigten Ausschnitt dieser Oper nicht sichtbar. MOrchester = musikalische Leistung des Orchesters, MSolisten = musikalische Leistung des Solisten, MChor = musikalische Leistung des Chores, MDirigent = Leistung des Dirigenten, SSolisten = szenische Leistung des Solisten, SChor = szenische Leistung des Chores, SAusstattung = Ausstattung (Einzelaspekte), fit3_Orchester = Zusammenspiel innerhalb des Orchesters, fit3_Chor = Zusammenspiel innerhalb des Chores, fit3_Solisten = Zusammenspiel der Solisten untereinander, fit3_Ausstattung = Stimmigkeit der Ausstattung, fit2_Szene = Stimmigkeit innerhalb der szenischen Dimension, fit2_Musik = Stimmigkeit innerhalb der musikalischen Dimension, fit1 = Stimmigkeit zwischen der musikalischen und der szenischen Dimension.

Anhang E: Messmodell der Teilstruktur II (Stimmigkeit innerhalb der szenischen Dimension)

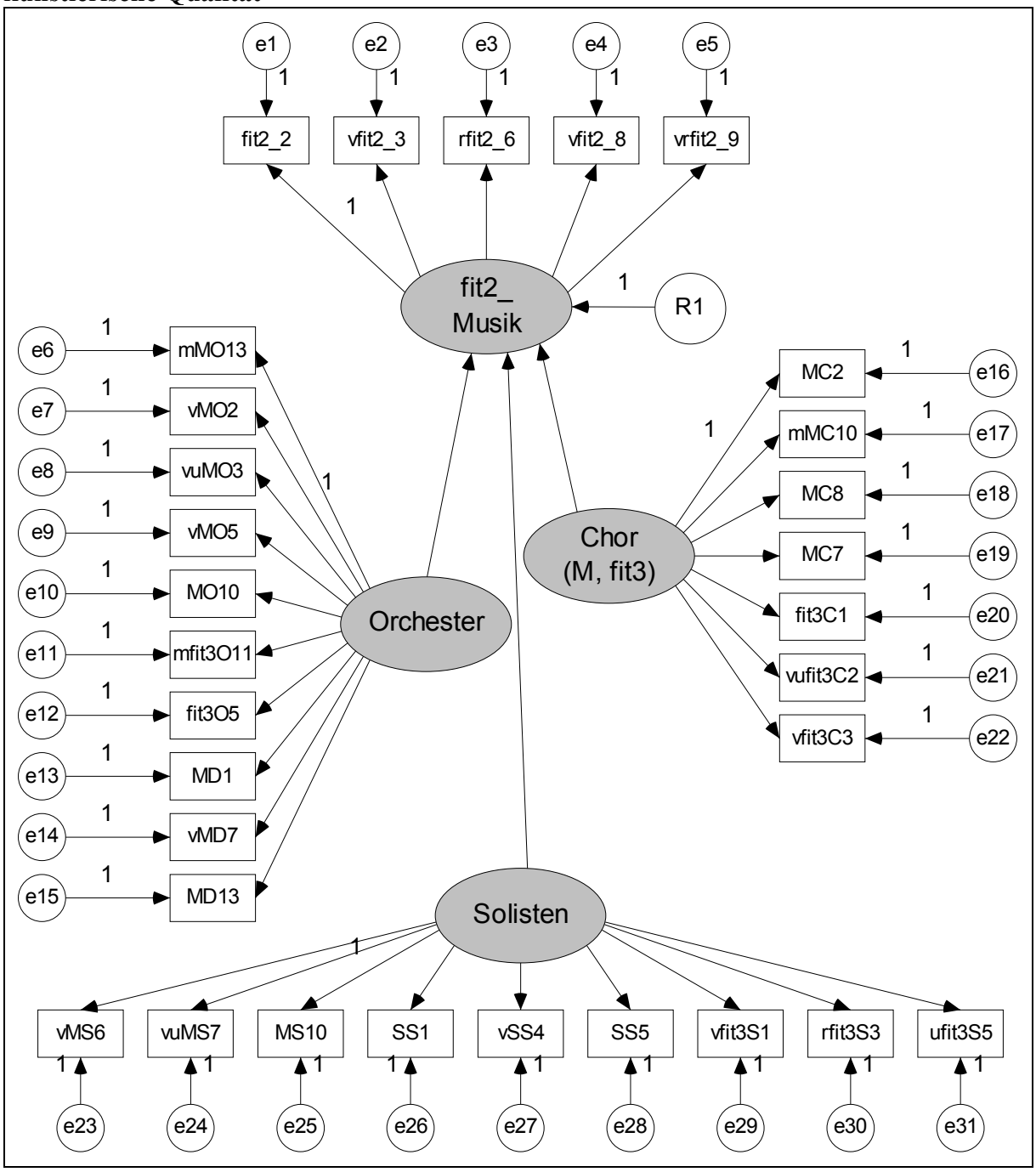
Abbildung 14: Teilstruktur II des Strukturgleichungsmodells des Urteils über die künstlerische Qualität



Anmerkungen: Grau schraffierte Ellipsen präsentieren latente Variablen, Rechtecke präsentieren manifeste Variablen, ‚e' stellen den jeweiligen Messfehler dar, ‚R' sind Residualvariablen. Werte am linken unteren Rand der latenten Variablen stellen die erklärte Varianz dar, Werte am rechten oberen bzw. am linken Rand der manifesten Variablen repräsentieren die Kommunalität. Pfeile mit Linien stellen Pfade dar, Werte bei den Pfeilen des Strukturmodells stellen die standardisierten ß-Koeffizienten dar, Werte bei den Pfeilen der Messmodelle repräsentieren die Faktorladungen. Es wurden die Kennziffern der jeweiligen Items verwendet; die Formulierungen können Anhang A entnommen werden.

Anhang F: Messmodell der Teilstruktur III (Stimmigkeit innerhalb der musikalischen Dimension)

Abbildung 15: Teilstruktur III des Strukturgleichungsmodells des Urteils über die künstlerische Qualität



Anmerkungen: Grau schraffierte Ellipsen präsentieren latente Variablen, Rechtecke präsentieren manifeste Variablen, ‚e' stellen den jeweiligen Messfehler dar, ‚R' sind Residualvariablen. Werte am linken unteren Rand der latenten Variablen stellen die erklärte Varianz dar, Werte am rechten oberen bzw. am linken Rand der manifesten Variablen repräsentieren die Kommunalität. Pfeile mit Linien stellen Pfade dar, Werte bei den Pfeilen des Strukturmodells stellen die standardisierten ß-Koeffizienten dar, Werte bei den Pfeilen der Messmodelle repräsentieren die Faktorladungen. Es wurden die Kennziffern der jeweiligen Items verwendet; die Formulierungen können Anhang A entnommen werden.

Anhang G: Wichtigkeit verschiedener Akteure für die künstlerische Qualität einer Opernaufführung im Allgemeinen

Tabelle 30: Wichtigkeit der verschiedenen Akteure für die künstlerische Qualität einer Opernaufführung – Laien

in % der gültigen Antworten	sehr wichtig	ziemlich wichtig	teils, teils	eher unwichtig	sehr unwichtig	Mittel-wert
Orchester	83.9	16.1	--	--	--	4.84
Solisten	98.2	1.8	--	--	--	4.98
Chor	44.6	35.7	17.9	1.8	--	4.23
Dirigent	75.0	23.2	1.8	--	--	4.73
Regisseur	78.6	17.9	3.6	--	--	4.75
Bühnenbildner	53.6	41.1	3.6	1.8	--	4.46
Techniker	44.6	48.2	7.1	--	--	4.38
Beleuchter	42.9	46.4	8.9	1.8		4.30

Tabelle 31: Wichtigkeit der verschiedenen Akteure für die künstlerische Qualität einer Opernaufführung – Experten

in % der gültigen Antworten	sehr wichtig	ziemlich wichtig	teils, teils	eher unwichtig	sehr unwichtig	Mittel-wert
Orchester	93.0	7.0	--	--	--	4.93
Solisten	91.2	8.8	--	--	--	4.91
Chor	71.9	22.8	3.5	1.8	--	4.65
Dirigent	96.5	3.5	--	--	--	4.96
Regisseur	77.2	17.5	5.3	--	--	4.72
Bühnenbildner	52.6	43.9	3.5	--	--	4.49
Techniker	52.6	42.1	3.5	1.8	--	4.46
Beleuchter	57.9	38.6	1.8	1.8	--	4.53

Anhang H: Faktorenanalysen zu Untersuchungsfrage 3

Tabelle 32: Faktorenanalyse der abhängigen Variablen ‚Gesamturteil'

Kennziffern der Items	Label der Items	Faktorladungen
mGE1	Erwartungen	.96
mGE2	Erlebnis	.96
mrGE3	enttäuscht	.89
mGE4	Erinnerung	.89
Varianzaufklärung		86%

Anmerkungen: Extraktionsmethode: Hauptkomponentenanalyse. Keine Vorgabe der Anzahl von Faktoren. Die Formulierungen der Items können Anhang A entnommen werden.

Tabelle 33: Faktorenanalyse der unabhängigen Variablen ‚künstlerische Qualität'

Kennziffern der Items	Label der Items	Faktorladungen
mGQ1	Überzeugung Qualität	.93
mGQ2	Spitzenleistung	.94
vNG2	Kritik	.88
Varianzaufklärung		84%

Anmerkungen: Extraktionsmethode: Hauptkomponentenanalyse. Keine Vorgabe der Anzahl von Faktoren. Die Formulierungen der Items können Anhang A entnommen werden.

Tabelle 34: Faktorenanalyse der unabhängigen Variablen ‚emotionale Wirkung'

Kennziffern der Items	Label der Items	Faktor 1	Faktor 2	Faktor 3
mEE1	Nachempfinden	.54	**.74**	.03
mEE2	Mitfiebern	.44	**.66**	.32
mEE3	Mitfühlen	.44	**.77**	.21
mEA1	positive Erinnerungen	.40	**.73**	.12
mEA3	eigenes Leben	.09	**.68**	.45
mEA4	Alltagsprobleme	-.15	.47	**.66**
mEP1	Genießen	**.85**	.29	.10
mEP2	Entspannung	**.81**	.31	.08
mEP3_2	Stimmung_während	**.76**	.28	.05
mEI1	wie im Flug	**.70**	.28	.24
vKA4_EI	Vergessen	**.71**	.38	.37
mEI3	Feuer und Flamme	**.69**	.37	.47
mES1	spezielle Momente	**.60**	.37	.43
mES2	physische Reaktion	.36	.12	**.78**
mES3	aufgewühlt	.30	.14	**.83**
rNG7	kalt gelassen	**.73**	.13	.31
vNG6	Berührung	.58	.18	**.62**
mrEM3	nicht angesprochen	**.75**	.17	.34
Varianzaufklärung		35%	20%	18%

Anmerkungen: Extraktionsmethode: Hauptkomponentenanalyse. Keine Vorgabe der Anzahl von Faktoren. Die Formulierungen der Items können Anhang A entnommen werden.

Tabelle 35: Faktorenanalyse der unabhängigen Variablen ‚Komplexität'

Kennziffern der Items	Label der Items	Faktorladungen
rKA1_k	hinschauen	.72
rKA2_k	Fortgang der Handlung	.84
rKA3_k	Bewegung der Darsteller	.86
Varianzaufklärung		66%

Anmerkungen: Extraktionsmethode: Hauptkomponentenanalyse. Keine Vorgabe der Anzahl von Faktoren. Die Formulierungen der Items können Anhang A entnommen werden.

Tabelle 36: Faktorenanalyse der unabhängigen Variablen ‚Vertrautheit'

Kennziffern der Items	Label der Items	Faktorladungen
rKA9_v	Art der Aufführung	.89
ruGA6_v	neuartig	.85
vrKA7_v	verwirrende Stilmittel	.91
Varianzaufklärung		78%

Anmerkungen: Extraktionsmethode: Hauptkomponentenanalyse. Keine Vorgabe der Anzahl von Faktoren. Die Formulierungen der Items können Anhang A entnommen werden.

Tabelle 37: Faktorenanalyse der unabhängigen Variablen ‚Identifikation'

Kennziffern der Items	Label der Items	Faktorladungen
mII1_i	Verhalten	.94
mII2_i	identifizieren	.93
mII3_i	Werte	.89
Varianzaufklärung		85%

Anmerkungen: Extraktionsmethode: Hauptkomponentenanalyse. Keine Vorgabe der Anzahl von Faktoren. Die Formulierungen der Items können Anhang A entnommen werden.

Tabelle 38: Faktorenanalyse der unabhängigen Variablen ‚situative Rahmenbedingungen'

Kennziffern der Items	Label der Items	Faktorladungen ‚Aufführungssaal'	Faktorladungen ‚Umfeld'
mRR1	Akustik	**.70**	.31
mRR2	Sitz	**.73**	-.01
mRR3	Sicht	**.66**	.29
mRR4	Ablenkungen	**.53**	.19
mRR5	Temperatur	**.61**	.02
mRR6	Service	.05	**.76**
mRR7	Räumlichkeiten	.12	**.89**
mRR8	Atmosphäre im Theater	.29	**.69**
Varianzaufklärung		28%	26%

Anmerkungen: Extraktionsmethode: Hauptkomponentenanalyse. Rotationsmethode: Varimax mit Kaiser-Normalisierung. Keine Vorgabe der Anzahl von Faktoren. Die Formulierungen der Items können Anhang A entnommen werden.

Anhang I: Informationsmaterial für die Teilnehmer der Befragung

Text, den sämtliche Zuschauer beim Betreten des Aufführungssaales überreicht bekamen

Besucherbefragung

Liebe Besucherin, lieber Besucher,

um künftig noch gezielter auf **Ihre Anregungen und Wünsche** einzugehen, führt die Universität Konstanz in Zusammenarbeit mit dem Anhaltischen Theater Dessau heute eine Besucherbefragung durch.

Dafür bitten wir um Ihre Mithilfe!

Den Fragebogen erhalten Sie während der Pause. Ideal für die Befragung wäre es, wenn Sie den Fragebogen **nach der Aufführung** beantworten und in die bereitgestellten Kästen am Ausgang geben. Wenn dies für Sie nicht möglich sein sollte, haben Sie die Möglichkeit, den Fragebogen bereits **in der Pause** auszufüllen. Es besteht ebenfalls die Möglichkeit, den Fragebogen **zu Hause** auszufüllen und ihn an uns zu senden. Einen vorbereiteten Freiumschlag erhalten Sie am Ausgang.

Als kleines Dankeschön für Ihre Unterstützung verlosen wir unter allen Teilnehmern **5 x zwei Theaterkarten** des Anhaltischen Theaters Dessau.

Vielen Dank für Ihre Mitarbeit!

Anhang J: Häufigkeiten der Antwortkategorien „habe ich nicht beachtet“ und „kann ich nicht beurteilen“ getrennt nach Laien und Experten

Tabelle 39: Häufigkeit der Antwortkategorien „habe ich nicht beachtet“ und „kann ich nicht beurteilen“ getrennt nach Laien und Experten

Kennziffer des Items	fehlend (in Prozent der gültigen Antworten)	davon (in % der Fehlenden) - Laien		davon (in % der Fehlenden) - Experten	
		„habe ich nicht beachtet“	„kann ich nicht beurteilen“	„habe ich nicht beachtet“	„kann ich nicht beurteilen“
mMO13	4.2	0.0	100.0	--	--
vMO2	7.5	37.5	62.5	--	--
vuMO3	1.7	0.0	100.0	--	--
vfit3O11	5.0	0.0	100.0	20.0	80.0
fit3O5	5.9	0.0	100.0	0.0	100.0
MD1	9.2	16.7	83.3	0.0	100.0
rMD4	7.6	0.0	100.0	0.0	100.0
vMD7	6.7	37.5	62.5	--	--
MD13	3.3	25.0	75.0	--	--
MS1	7.6	42.9	57.1	100.0	0.0
vrMS2	2.5	0.0	100.0	0.0	100.0
MS5	0.8	--	--	0.0	100.0
vMS6	4.2	66.7	33.2	50.0	50.0
vuMS7	0.8	0.0	100.0	--	--
MS10	1.7	0.0	100.0	--	--
MC2	5.0	33.3	66.7	--	--
mMC10	5.0	50.0	50.0	--	--
MC8	5.0	16.7	83.3	--	--
MC7	5.0	16.7	83.3	--	--
SS1	0.8	0.0	100.0	--	--
vSS4	0.8	0.0	100.0	--	--
vSS8	0.8	0.0	100.0	--	--
rSS10	2.5	0.0	100.0	0.0	100.0
vuSC1	1.6	50.0	50.0	--	--
vSC2	1.7	0.0	100.0	--	--
vSC3	2.5	100.0	0.0	--	--
vuSC4	3.3	25.0	75.0	--	--
SC6	5.0	83.3	16.7	--	--
fit3C1	3.4	0.0	100.0	--	--
vufit3C2	9.4	50.0	50.0	100.0	0.0
vfit3C3	3.4	100.0	0.0	100.0	0.0
vfit3S1	0.8	0.0	100.0	--	--

(Fortsetzung) Kennziffer des Items	fehlend (in Prozent der gültigen Antworten)	davon (in % der Fehlenden) - Laien		davon (in % der Fehlenden) - Experten	
		„habe ich nicht beachtet“	„kann ich nicht beurteilen“	„habe ich nicht beachtet“	„kann ich nicht beurteilen“
rfit3S3	3.3	0.0	100.0	--	--
ufit3S5	1.7	0.0	100.0	--	--
urfit3S6	3.3	0.0	100.0	--	--
fit2_2	0.8	0.0	100.0	--	--
rfit2_6	4.2	66.7	33.2	100.0	0.0
vrfit2_9	2.5	0.0	100.0	--	--
rfit3A3	2.5	50.0	50.0	--	--
fit3A5	0.8	100.0	0.0	--	--
vrfit2S_4	1.7	0.0	100.0	--	--
vfit2S_3	0.9	0.0	100.0	--	--
vfit1_1	6.0	20.0	80.0	0.0	100.0
rfit1_2	3.5	0.0	100.0	0.0	100.0
fit1_3	3.4	0.0	100.0	0.0	100.0
fit1_6	2.6	0.0	100.0	0.0	100.0
fit1_7	2.6	0.0	100.0	0.0	100.0
vfit1_8	2.6	0.0	100.0	0.0	100.0

Anmerkung: Die Formulierung der Items kann Anhang A entnommen werden.

Anhang K: Prozentuale Häufigkeit nicht beantworteter Items getrennt nach Laien und Experten

Anmerkung: Fett markiert sind die Items, die von jeweils mindestens 5% der Befragten nicht beantwortet werden konnten.

Tabelle 40: Prozentuale Häufigkeit nicht beantworteter Items getrennt nach Laien und Experten

Kennziffer des Items	fehlend (in Prozent aller gültigen Antworten)	Laien	Exper-ten
Musikalische Leistung und Zusammenspiel des Orchesters sowie Leistung des Dirigenten			
mMO13	4.2	**8.9**	0.0
vMO2	**7.5**	**8.9**	0.0
vuMO3	1.7	3.5	0.0
vMO5	--	--	--
MO10	--	--	--
vfit3O11	**5.0**	1.8	**8.8**
fit3O5	**5.9**	**5.4**	**7.0**
MD1	**9.2**	**10.7**	**8.8**
rMD4	**7.6**	**5.4**	**8.8**
vMD7	**6.7**	**14.3**	0.0
MD13	3.3	**7.2**	0.0
Musikalische Leistung des Solisten			
MS1	**7.6**	**12.5**	3.5
vrMS2	2.5	1.8	1.8
MS5	0.8	0.0	1.8
vMS6	4.2	**5.3**	3.6
vuMS7	0.8	1.8	0.0
MS10	1.7	3.5	0.0
Musikalische Leistung des Chores			
MC2	**5.0**	**10.5**	0.0
mMC10	**5.0**	**10.6**	0.0
MC8	**5.0**	**10.6**	0.0
MC7	**5.0**	**10.6**	0.0
Szenische Leistung des Solisten			
SS1	0.8	1.8	0.0
vSS4	0.8	1.8	0.0

(Fortsetzung) Kennziffer des Items	fehlend (in Prozent aller gültigen Antworten)	Laien	Exper-ten
Szenische Leistung des Solisten			
SS5	--	--	--
vSS8	0.8	1.8	0.0
rSS10	2.5	1.8	1.8
Szenische Leistung des Chores			
vuSC1	1.6	3.6	0.0
vSC2	1.7	3.6	0.0
vSC3	2.5	**5.4**	0.0
vuSC4	3.3	**7.2**	0.0
SC6	**5.0**	**10.7**	0.0
Zusammenspiel innerhalb des Chores			
fit3C1	3.4	**7.1**	0.0
vufit3C2	**9.4**	**17.8**	1.8
vfit3C3	3.4	**5.4**	1.8
Zusammenspiel aller Solisten untereinander			
vfit3S1	0.8	1.8	0.0
rfit3S3	3.3	**5.3**	0.0
ufit3S5	1.7	3.5	0.0
urfit3S6	3.3	3.3	0.0
Zusammenspiel von Orchester, Solisten und Chor			
fit2_2	0.8	1.8	0.0
vfit2_3	--	--	--
rfit2_6	4.2	**5.3**	1.8
vfit2_8	--	--	--
vrfit2_9	2.5	3.5	0.0
Stimmigkeit der szenischen Aspekte			
fit3A1	--	--	--
vfit2S_1	--	--	--
fit2S_2	--	--	--
vfit3A2	--	--	--
fit3A4	--	--	--
rfit3A3	2.5	3.6	0.0
fit3A5	0.8	1.8	0.0

(Fortsetzung) Kennziffer des Items	fehlend (in Prozent aller gültigen Antworten)	Laien	Exper-ten
Stimmigkeit der szenischen Aspekte			
vrfit2S_4	1.7	1.9	0.0
vfit2S_3	0.9	1.8	0.0
Stimmigkeit zwischen Musik und Szene			
vfit1_1	**6.0**	**8.9**	3.6
rfit1_2	3.5	3.6	1.8
fit1_3	3.4	3.6	3.6
fit1_6	2.6	3.6	1.8
fit1_7	2.6	1.8	3.8
vfit1_8	2.6	1.8	3.7

Anmerkung: Die Formulierungen der Items können Anhang A entnommen werden.

Anhang L: Interkorrelation der ausgewählter Skalen

Tabelle 41: Interkorrelation der Skalen ‚musikalische Leistung des Orchesters', ‚Zusammenspiel innerhalb des Orchesters' und ‚Leistung des Dirigenten'

	Orchester	MOrchester	fit3_Orchester	Dirigent
Orchester	1			
MOrchester	.98**	1		
fit3_Orchester	.95**	.89**	1	
Dirigent	.96**	.90**	.90**	1

Anmerkungen: **p≤ 0.01; n = 91 (listenweiser Fallausschluss). Orchester = zusammengefasste Skala (musikalische Leistung und Zusammenspiel des Orchesters sowie Leistung des Dirigenten), MOrchester = musikalische Leistung des Orchesters, fit3_Orchester = Zusammenspiel innerhalb des Orchesters, Dirigent = Leistung des Dirigenten.

Tabelle 42: Interkorrelation der Skalen ‚Stimmigkeit innerhalb der szenischen Dimension' und ‚Stimmigkeit der Ausstattung'

	fit2_Szene	fit2_Szene (solo)	fit3_Ausstattung
fit2_Szene	1		
fit2_Szene (solo)	.98**	1	
fit3_Ausstattung	.95**	.87**	1

Anmerkungen: **p≤ 0.01; n = 111 (listenweiser Fallausschluss). fit2_Szene = zusammengefasste Skala (Stimmigkeit innerhalb der szenischen Dimension und Stimmigkeit der szenischen Aspekte), fit2_Szene (solo) = Stimmigkeit innerhalb der szenischen Dimension, fit3_Ausstattung = Stimmigkeit der szenischen Aspekte.

Anhang M: Faktorenanalysen getrennt nach Laien und Experten

Tabelle 43: Faktorenanalyse der ersten Ebene - Laien

Kennziffern der Items	Label der Items	fit1 Faktorladungen
vfit1_1	Musik – Szene	.88
fit1_3	Charakter der Musik	.87
fit1_6	unterstrichen	.78
fit1_7	Musik und Szene wie aus einem Guss	.92
vfit1_8	Unterstützung	.90
Varianzaufklärung		76%

Anmerkungen: Extraktionsmethode: Hauptkomponentenanalyse. Keine Vorgabe der Anzahl von Faktoren. Die Formulierungen der Items können Anhang A entnommen werden.

Tabelle 44: Faktorenanalyse der ersten Ebene - Experten

Kennziffern der Items	Label der Items	fit1 Faktorladungen
vfit1_1	Musik – Szene	.93
fit1_3	Charakter der Musik	.91
fit1_6	unterstrichen	.83
fit1_7	Musik und Szene wie aus einem Guss	.88
vfit1_8	Unterstützung	.93
Varianzaufklärung		81%

Anmerkungen: Extraktionsmethode: Hauptkomponentenanalyse. Keine Vorgabe der Anzahl von Faktoren. Die Formulierungen der Items können Anhang A entnommen werden.

Tabelle 45: Faktorenanalyse der zweiten Ebene – Laien

Kennziffern der Items	Label der Items	fit2_Szene Faktorladungen	fit2_Musik Faktorladungen
fit2_2	Stahlkraft und Volumen	.04	**.77**
vfit2_3	Abstimmung Lautstärke	.18	**.83**
rfit2_6	übertönen	.23	**.77**
vfit2_8	Zudecken	.18	**.77**
vrfit2_9	Überflutung	-.01	**.76**
fit3A1	Atmosphäre Bühnenbild	**.81**	.23
vfit2S_1	Handlung und Darsteller	**.88**	.14
fit2S_2	Dekoration	**.86**	.08
vfit3A2	Ausstattung	**.92**	.05
fit3A4	Abstimmung Kostüme - Bühnenbild	**.79**	.19
fit3A5	Requisiten und Licht	**.87**	.01
vrfit2S_4	Fremdkörper	**.80**	.13
vfit2S_3	stimmiges Bild	**.85**	.19
Varianzaufklärung		45%	25%

Anmerkungen: Extraktionsmethode: Hauptkomponentenanalyse. Rotationsmethode: Varimax mit Kaiser-Normalisierung. Vorgabe von zwei Faktoren. Die Formulierungen der Items können Anhang A entnommen werden.

Tabelle 46: Faktorenanalyse der zweiten Ebene – Experten

Kennziffern der Items	Label der Items	fit2_Szene Faktorladungen	fit2_Musik Faktorladungen
fit2_2	Stahlkraft und Volumen	.21	**.84**
vfit2_3	Abstimmung Lautstärke	.36	**.82**
rfit2_6	übertönen	.05	**.71**
vfit2_8	Zudecken	.19	**.78**
vrfit2_9	Überflutung	.04	**.62**
fit3A1	Atmosphäre Bühnenbild	**.88**	-.02
vfit2S_1	Handlung und Darsteller	**.91**	-.06
fit2S_2	Dekoration	**.87**	.18
vfit3A2	Ausstattung	**.85**	.30
fit3A4	Abstimmung Kostüme - Bühnenbild	**.86**	.33
fit3A5	Requisiten und Licht	**.86**	.20
vrfit2S_4	Fremdkörper	**.77**	.23
vfit2S_3	stimmiges Bild	**.81**	.40
Varianzaufklärung		46%	26%

Anmerkungen: Extraktionsmethode: Hauptkomponentenanalyse. Rotationsmethode: Varimax mit Kaiser-Normalisierung. Vorgabe von zwei Faktoren. Die Formulierungen der Items können Anhang A entnommen werden.

Tabelle 47: Faktorenanalyse der dritten Ebene - Laien

Kennziffern der Items	Label der Items	Faktor 1	Faktor 2	Faktor 3	Faktor 4	Faktor 5	Faktor 6	Faktor 7
mMO13	gute Vorbereitung	**.69**	.00	.36	.19	.17	.46	.03
vMO2	Lautstärke Orchester	**.50**	-.01	.29	.08	.32	.49	.21
vuMO3	musikalische Spannung	**.95**	.09	.08	.11	.05	.09	-.03
vMO5	Spannung und Intensität	**.91**	.08	.07	.22	.18	.02	.00
MO10	Lebendigkeit	**.95**	.10	.14	.11	-.04	.07	.02
mfit3O11	Instrumentengruppen	**.87**	.07	.18	.17	.18	.14	.01
fit3O5	Orchester wie aus einem Guss	**.85**	.06	.14	.22	.24	.05	.15
MD1	im Griff	**.89**	.04	.19	.17	.06	-.10	.02
vMD7	Balance	**.64**	-.05	.27	.20	.28	.41	.22
MD13	dem Werk gewachsen	**.92**	.09	.24	.07	-.03	.18	.03
vMS6	Unterschiede in der Lautstärke	.33	-.15	-.04	**.83**	.23	.00	.16
vuMS7	der Partie gewachsen	.17	-.22	-.08	**.88**	.13	.09	.08
MS10	Anspruch der Rolle	.21	-.18	-.08	**.85**	.13	.14	-.16
MC2	zur Geltung	.25	-.01	**.92**	-.06	.03	.06	.15
mMC10	musikalische Leistung	.37	.05	**.84**	.02	.05	.22	.04

(Fortsetzung) Kennziffern der Items	Label der Items	Faktor 1	Faktor 2	Faktor 3	Faktor 4	Faktor 5	Faktor 6	Faktor 7
MC8	in jeder Hinsicht	.05	-.03	**.93**	.05	.16	-.04	.20
MC7	klangliche Brillanz	.29	-.09	**.88**	-.12	-.01	.10	-.09
SS1	eindringlich	.47	.41	.05	**.54**	.06	-.28	.20
vSS4	Leben der Rolle	.10	.50	.05	.13	-.14	-.09	**.77**
SS5	Verkörperung	.56	.31	.07	**.58**	-.24	-.08	.10
vuSC1	Erscheinungsbild	.10	**.94**	-.05	-.01	.06	-.11	-.02
vSC2	voll abnehmen	-.01	**.90**	-.04	.02	-.01	-.18	.13
vSC3	ästhetisch befriedigend	.00	**.91**	-.10	-.04	.22	-.14	-.09
vSA13	Bühnenbild gefallen	.18	**.81**	-.01	-.16	.01	.27	.05
mSA16	Bühnenbild beeindruckend	.13	**.73**	.03	-.23	.13	.46	.09
vSA15	Kostüme gefallen	.06	**.72**	.16	-.08	.14	.50	.01
fit3C1	Homogenität	.18	.30	-.01	.15	**.86**	.01	-.09
vufit3C2	Tempo	.26	.10	.25	.15	**.77**	.13	.27
vfit3C3	Stimmen	.09	-.17	.34	-.05	.27	.18	**.71**
vfit3S1	Abstimmung Solisten	**.47**	.23	.18	.39	.44	.41	-.06
rfit3S3	Solisten gegeneinander	.18	.03	.12	.52	-.15	**.54**	-.09
ufit3S5	Harmonie	**.57**	-.04	.18	.23	.23	**.51**	.25
Varianzaufklärung		27%	16%	12%	12%	7%	7%	5%

Anmerkungen: Extraktionsmethode: Hauptkomponentenanalyse. Rotationsmethode: Varimax mit Kaiser-Normalisierung. Ohne Vorgabe der Anzahl von Faktoren. Die Formulierungen der Items können Anhang A entnommen werden.

Tabelle 48: Faktorenanalyse der dritten Ebene – Experten

Kennziffern der Items	Label der Items	Faktor 1	Faktor 2	Faktor 3	Faktor 4	Faktor 5	Faktor 6
mMO13	gute Vorbereitung	**.74**	-.03	.31	-.09	.30	.38
vMO2	Lautstärke Orchester	**.81**	.32	.09	.22	.07	.17
vuMO3	musikalische Spannung	**.88**	.23	.18	.09	.08	.15
vMO5	Spannung und Intensität	**.83**	.25	.21	.19	.18	.12
MO10	Lebendigkeit	**.84**	.28	.15	.21	.12	.19
mfit3O11	Instrumentengruppen	**.86**	.34	.04	.15	.13	.00
fit3O5	Orchester wie aus einem Guss	**.82**	.25	.07	.23	.29	.02
MD1	im Griff	**.79**	.39	.14	.05	.13	.08

(Fortsetzung) Kennziffern der Items	Label der Items	Faktor 1	Faktor 2	Faktor 3	Faktor 4	Faktor 5	Faktor 6
vMD7	Balance	**.63**	.34	-.01	.45	.33	.04
MD13	dem Werk gewachsen	**.80**	.40	.10	.14	.23	.08
vMS6	Unterschiede in der Lautstärke	.32	**.58**	.18	.45	.18	.37
vuMS7	der Partie gewachsen	.15	.40	.17	.25	.16	**.78**
MS10	Anspruch der Rolle	.21	.46	.19	.28	.24	**.70**
MC2	zur Geltung	.32	**.82**	.26	.06	.17	.19
mMC10	musikalische Leistung	.33	**.84**	.29	.08	.12	.11
MC8	in jeder Hinsicht	.38	**.81**	.27	.13	.01	.12
MC7	klangliche Brillanz	.30	**.80**	.26	.02	.10	.24
SS1	eindringlich	.49	**.60**	.27	.34	.13	.24
vSS4	Leben der Rolle	.40	**.66**	.22	.27	.08	.16
SS5	Verkörperung	.46	**.48**	.25	.21	.23	.48
vuSC1	Erscheinungsbild	.14	.26	**.89**	.12	.14	.15
vSC2	voll abnehmen	.16	.37	**.82**	.10	.09	.24
vSC3	ästhetisch befriedigend	.17	.30	**.88**	.12	-.02	.05
vSA13	Bühnenbild gefallen	.21	.07	.11	**.82**	.15	.25
mSA16	Bühnenbild beeindruckend	.20	.15	.11	**.88**	.19	.18
vSA15	Kostüme gefallen	.19	.31	.51	**.56**	.09	-.11
fit3C1	Homogenität	.35	**.67**	.34	.11	.37	.17
vufit3C2	Tempo	.27	**.60**	.16	.34	.51	.04
vfit3C3	Stimmen	.25	.09	.07	.12	**.82**	.15
vfit3S1	Abstimmung Solisten	**.53**	.07	.02	.38	.48	.42
rfit3S3	Solisten gegeneinander	.38	.04	.19	.15	**.65**	.47
ufit3S5	Harmonie	.18	.44	.01	.23	**.68**	.03
Varianzaufklärung		27%	21%	11%	10%	9%	8%

Anmerkungen: Extraktionsmethode: Hauptkomponentenanalyse. Rotationsmethode: Varimax mit Kaiser-Normalisierung. Ohne Vorgabe der Anzahl von Faktoren. Die Formulierungen der Items können Anhang A entnommen werden.

Literaturverzeichnis

Adorno, Th. W. (1973): Studien zum autoritären Charakter. Frankfurt a.M.: Suhrkamp.

Allmann, U. (1997): Innovatives Theatermanagement. Eine Fallstudie. Wiesbaden: Deutscher Universitätsverlag.

Aristoteles (in der Übersetzung von Lucas, D.) (1968): Poetics. Oxford: Clarendon.

Backhaus, K., Erichson, B., Plinke, W. und Weiber, R. (2003): Multivariate Analysemethoden. Eine anwendungsorientierte Einführung. Berlin: Springer Verlag.

Baruch, Y. und Ramalho, N. (2006): Communalities and Distinctions in the Measurement of Organizational Performance and Effectiveness Across For-Profit and Nonprofit Sectors. *Nonprofit and Voluntary Sector Quarterly*. 35(1), S. 39 – 65.

Baumol, W. und Bowen, W. (1966): Performing Arts – The Economic Dilemma. A Study of Problems Common to Theatre, Opera, Music, and Dance. New York: Twentieth Cent. Fund.

Behne. K. (1987): Urteile und Vorurteile. Die Alltagsmusiktheorien jugendlicher Hörer. S. 186 – 220. In: de la Motte-Haber (Hg.): Psychologische Grundlagen des Musiklernens. Kassel: Bärenreiter.

Behne, K. (1997): Musikästhetik. III. Musikalische Urteilsbildung. S. 998 – 1016. In: L. Finscher (Hg.): Die Musik in Geschichte und Gegenwart. Sachteil Bd. 6. Kassel: Bärenreiter/Metzler.

Behr, M. (1983): Musiktheater. Faszination, Wirkung, Funktion. Wilhelmshaven: Heinrichshofen.

Berlyne, D. (1960): Conflict, Arousal, and Curiosity. New York: McGraw-Hill.

Berlyne, D. (1970): Novelty, Complexity, and Hedonic Value. *Perception and Psychophysics*. 8, S. 279 – 286.

Berlyne, D. (1971): Aesthetics and Psychobiology. New York: Appleton.

Berlyne, D. (1974a): Konflikt, Erregung, Neugier. Zur Psychologie der kognitiven Motivation. Stuttgart: Ernst Klett.

Berlyne, D. (1974b): Studies in the New Experimental Aesthetics. Steps toward an Objective Psychology of Aesthetic Appreciation. Washington, DC: Hemisphere.

Bersch-Burauel, A. (2004): Entwicklung von Musikpräferenzen im Erwachsenenalter. Eine explorative Untersuchung. Unveröffentlichte Dissertation, Universität Paderborn.

Boerner, S. (2002): Führungsverhalten und Führungserfolg. Beitrag zu einer Theorie der Führung am Beispiel des Musiktheaters. Wiesbaden: Deutscher Universitätsverlag.

Boerner, S. und Renz, S. (in Vorbereitung): Performance Measurement in Theatre Companies. Comparing Subjective Quality Judgments of Experts and Non Experts in Opera.

Boerner, S., Neuhoff, H., Renz, S. und Moser, V. (in Druck): Evaluation in Music Theatre: Empirical Results on Content and Structure of the Audience's Quality Judgment. *Empirical Studies of the Arts*. 26(1).

Bogumil, J. und Naschold, F. (2000): Modernisierung des Staates. New Public Management in deutscher und internationaler Perspektive. Opladen: Leske + Budrich.

Boorsma, M. und van Maanen, H. (2003): View and review in the Netherlands: The Role of Theatre Critics in the Construction of Audience Experience. *International Journal of Cultural Policy*. 9(3), S. 319 – 335.

Bortz, J. (1999): Statistik für Sozialwissenschaftler. Berlin: Springer Verlag.

Boyle, J. und Radcoy, R. (1987): Measurement and Evaluation of Musical Experience. New York: Schirmer Books.

Braun, W. (1972): Musikkritik. Versuch einer historisch-kritischen Standortbestimmung. Köln: Gerig.

Brosius, F. (2006): SPSS 14. Heidelberg: mitp.

Browne, M. und Cudeck, R. (1993): Alternative Ways of Assessing Model Fit. S. 136 - 162. In: Bollen, K. und Long, J. (Hg.): Testing structural equation models. Newbury Park, CA: Sage.

Butzer-Strothmann, K., Günter, B. und Degen, H. (2001): Leitfaden für Besucherbefragungen durch Theater und Orchester. Baden-Baden: Nomos.

Byo, J. und Brooks, R. (1994): A Comparison of Junior High Musicians' and Music Educators' Performance Evaluations of Instrumental Music. *Contributions to Music Education*. 21, S. 26 – 38.

Chaffee, E. und Sherr, L. (1993): Total Quality Management. S. 3 – 23. In: Connors, D. (Hg.): The Nonprofit Management Handbook: Operating Policies and Procedures. New York: Wiley.

Chi, M., Glaser, R. und Farr, M. (1988): The Nature of Expertise. Hillsdale, NJ: Erlbaum.

Csikszentmihalyi, M. (1991): Flow. The Psychology of Optimal Experience. New York: Harper & Row.

Csikszentmihalyi, M. und Robinson, R. (1990): The Art of Seeing. An Interpretation of the Aesthetic Encounter. Los Angeles: J. Paul Getty Museum Publications.

de la Motte-Haber, H. (1973): Der Einfluss psychologischer Variablen auf das ästhetische Urteil. S. 163 – 174. In: Jahrbuch des Staatlichen Instituts für Musikforschung 1972/73. Preußischer Kulturbesitz. Berlin : De Gruyter.

Deutscher Bühnenverein: Theaterstatistik 2004/05. URL: http://www.buehnenverein.de/presse/statistik_thstatistik.php. (18.06.2007).

Deutscher Bühnenverein (2006): Wer spielte was? Werkstatistik 2004/05 des Deutschen Bühnenvereins. Darmstadt: Mykenae Verlag.

Dillman, D. (1978): Mail and Telephone Surveys. The Total Design Method. New York: Wiley.

Dillman, D. (2007): Mail and Internet Surveys. The Tailored Design Method. New York: Wiley.

Dollase, R., Rüsenberg, M. und Stollenwerk, H. (1986): Demoskopie im Konzertsaal. Mainz: Schott.

Erhardt, M. (1994): Das soll das Theater! Eine vergleichende Untersuchung über Maßstäbe und Meinungen zum deutschen Theater der Gegenwart. Unveröffentlichte Magisterarbeit, LMU München.

Ericsson, K. und Smith, J. (1991): Towards a General Theory of Expertise: Prospects and Limits. Cambridge: Cambridge University Press.

Ericsson, K., Krampe, R. und Tesch-Römer, C. (1993): The Role of Deliberate Practice in the Acquisition of Expert Performance. *Psychological Review*. 100(3), S. 363-406.

Eversmann, P. (2004): The Experience of the Theatrical Event. S. 139 – 174. In: V. Cremona, P. Eversmann, H. van Maanen, W. Sauter und J. Tulloch (Hg.): Theatrical Evens – Borders, Dynamics, Frames. Amsterdam: Rodopi.

Fabel, M. (1998): Kulturpolitisches Controlling. Ziele, Instrumente und Prozesse der Theaterförderung in Berlin. Frankfurt a.M.: Lang.

Fischer, D., Hasse, M. und Biethan, J. (1994): Eine empirische Untersuchung zum Stand des Controlling und des Einsatzes von Controlling-Informationssystemen in Theatern und Opernhäusern. In: Biethan, J. und Schumann (Hg.): Göttinger Wirtschaftsinformatik. Arbeitsbericht Nr. 4.

Furman, C. und Duke, R. (1988): Effect of Majority Consensus on Preferences for Recorded Orchestral and Popular Music. *Journal of Research in Music Education*. 36(4), S. 220 – 231.

Gabrielsson, A. (2001): Emotions in Strong Experiences with Music. S. 431 – 449. In: Juslin, P. und Sloboda, J. (Hg.): Music and Emotion. Oxford: Oxford University Press.

Gembris, H. (1996): Rezeptionsforschung. S. 321 – 326. In: Decker-Voigt, H., Knill, P. und Weymann, E. (Hg.): Lexikon Musiktherapie. Göttingen: Hogrefe.

Göschel, A. (1997): Konzeption von Kulturpolitik und ihre Folgen für Kulturmanagement. S. 137 – 160. In: Heinrichs, W. (Hg.): Macht Kultur Gewinn? Kulturbetrieb zwischen Nutzen und Profit. Baden-Baden: Nomos.

Goldstein, A. (1980): Thrills in Response to Music and Other Stimuli. *Physiological Psychology*. 8, S. 126 – 129.

Hargreaves, D. (1986): The Developmental Psychology of Music. Cambridge: Cambridge University Press.

Hargreaves, D. und Colman, A. (1981): The Dimensions of Aesthetic Reactions to Music. *Psychology of Music.* 9, S. 15 – 20.

Harrer, G. und Allesch, C. (1982): Grundlagen der Musiktherapie und Musikpsychologie. Stuttgart: Fischer.

Harrer, G. (1993): Beziehung zwischen Musikwahrnehmung und Emotionen. S. 588 – 599. In: Bruhn, H., Oerter, R. und Rösing, H. (Hg.): Musikpsychologie. Ein Handbuch. Reinbek bei Hamburg.

Heinrichs, W. (1995): Privatisierung öffentlicher Kulturbetriebe aus kulturpolitischer Sicht. S. 296 – 308. In: Heinze, T. (Hg.): Kultur und Wirtschaft. Perspektiven gemeinsamer Innovation. Opladen: Westdeutscher Verlag.

Heinze, T. (1994): Einführung. S. 7 – 20. In: Heinze, T. (Hg.): Kulturmanagement. Professionalisierung kommunaler Kulturarbeit. Opladen: Westdeutscher Verlag.

Heinze, T. (1997): Kulturmanagement: Eine Annäherung. S. 48 – 75. In: Heinze, T. (Hg.): Kulturmanagement II. Konzepte und Strategien. Opladen: Westdeutscher Verlag.

Hoegl, C. (1995): Ökonomie der Oper. Grundlagen für das Musiktheater-Management. Bonn: ARCult Media.

Homburg, C. und Baumgartner, H. (1995): Beurteilung von Kausalmodellen. *Marketing ZFP*. 17, S. 162 – 176.

Hülshoff, T. (1999): Emotionen. München, Basel: Reinhardt.

Jungheinrich, H. (1986): Der Musikdarsteller. Zur Kunst des Dirigenten. Frankfurt a.M.: Fischer.

Koebner, T. (1993): Handlungen mit Musik. Die Oper als Zeitspiegel, Leidenschaftsdrama, Gesamtkunstwerk. Anif/Salzburg: Müller-Speiser.

Kommunale Gemeinschaftsstelle für Verwaltungsvereinfachung (KGSt) (1989): Führung und Steuerung des Theaters. Köln.

Konijn, E. (1999): Spotlight on Spectators: Emotions in the Theatre. *Discourse Processes*. 28(2), S. 169 – 194.

Kreitler, H. und Kreitler, S. (1980): Psychologie der Kunst. Stuttgart: Kohlhammer.

Kühn, H. (1982): Musiktheater heute. Eine Einführung. S. 9 – 19. In: Kühn, H. (Hg.): Musiktheater heute. Mainz: Schott.

Laski, M. (1961): Ecstasy. A Study of Some Secular and Religious Experiences. London: Cresset Press.

Lehmann, A. (1994): Habituelle und situative Rezeptionsweisen beim Musikhören. Eine einstellungstheoretischen Untersuchung. Frankfurt a.M.: Lang.

Lesle, L. (1981): Notfall Musikkritik. Wiesbaden: Breitkopf & Härtel.

Lipp, W. (1994): Drama Kultur. Berlin: Duncker & Humblot.

Maslow, A. (1968): Toward a Psychology of Being. New York: Van Nostrand Reinhold.

Maslow, A. (1976): The Farther Reaches of Humane Nature. New York: Penguin Books.

Mauser, S. (2002): Musiktheater im 20. Jahrhundert. Laaber: Laaber-Verlag.

McDougall, W. (1908): An Introduction to Social Psychology. London: Methuen.

McNamara, L. und Ballard, M. (1999): Resting Arousal, Sensation Seeking and Music Preference. *Genetic, Social and General Psychology Monographs*. 125(3), S. 229 – 250.

Morguet, M. und Moser-Hauck, J. (1991): Beurteilung von Musik. Gibt es Unterschiede zwischen Musikexperten und Laien? *Musikpädagogische Forschung.* 12, S. 189 – 199.

Moser, V. (2005): Qualitätsmessung in öffentlichen Dienstleistungsorganisationen. Die Entwicklung eines Instruments zur Evaluierung von Musiktheateraufführungen. Unveröffentlichte Diplomarbeit, Universität Konstanz.

Neuhoff, H. (2001): Die Altersstruktur von Konzertpublika. Querschnitte und Längsschnitte von Klassik und Pop in kultursoziologischer Analyse. S. 64 – 83. In: Deutscher Musikrat (Hg.): Musikforum 95.

Neuhoff, H. und de la Motte-Haber, H. (2007): Musikalische Sozialisation. S. 389 – 417. In: de la Motte-Haber, H. und Neuhoff, H.: Handbuch Musiksoziologie. Laaber: Laaber-Verlag.

Niketta, R. (1993): Urteils- und Meinungsbildung. S. 329 – 339. In: Bruhn, H., Oerter, R. und Rösing, H. (Hg.): Musikpsychologie. Ein Handbuch. Reinbek bei Hamburg: Rowohlt.

Nolte, F., Schäfer, T. und Yesilcicek, Ö. (2001) Opernpublikum – Musicalpublikum. Eine Studie zur Soziologie des Musiktheaters. URL: http://freenet-homepage.de/Frank.Nolte/Gesamt.PDF. (12.09.2007)

Ossadnik, W. (1987): Theatermanagement mittels Controlling. *Zeitschrift für öffentliche und gemeinwirtschaftliche Unternehmen.* 10(2), S. 145 – 157.

Pahlen, K. (1981): Oper der Welt. Zürich: SV International/ Schweizer Verlagshaus.

Panksepp, J. (1995): The Emotional Sources of ‘Chills’ Induced by Music. *Music Perception*. 13, S. 171 – 207.

Panzarella, R. (1980): The Phenomenology of Aesthetical Peak Experiences. *Journal of Humanistic Psychology*. 20, S. 69 – 85.

Pearl, J. (2000): Causality. Models, Reasoning, and Inference. Cambridge: Cambridge University Press.

Piaget, J. (1936): La Naissance de l'intelligence chez l'enfant. Neuchatel: Delachaux & Niestlé.

Pike, A. (1972): A Phenomenological Analysis of Emotional Experience in Music. *Journal of Research in Music Education*. 20, S. 262 – 267.

Raab, E. (1981): Ästhetik und Neugier. S. 263 – 308. In: Voss, H. und Keller, H. (Hg.): Neugier-Forschung. Grundlagen – Theorien – Anwendungen. Weinheim und Basel: Beltz Verlag.

Rawlings, D., Twomey, F., Burns, E. und Morris, S. (1998): Personality, Creativity, and Aesthetic Preference: Comparing Psychoticism, Sensation Seeking, Schizotypy, and Openness to Experience. *Empirical Studies of the Arts*. 16(2), S. 153 – 178.

Renz, S. (2006): Evaluation in Nonprofit Organisationen. Eine empirische Untersuchung zur Rezeption der Aufführungsqualität im Musiktheater. Unveröffentlichte Diplomarbeit, Universität Konstanz.

Reuband, K. (2005): Sterben die Opernbesucher aus? Eine Untersuchung zur sozialen Zusammensetzung des Opernpublikums im Zeitvergleich. S. 123 – 138. In: Klein, A. und Knubben, T. (Hg.): Deutsches Jahrbuch für Kulturmanagement. Baden-Baden: Nomos.

Roth, P. (1994): Missing Data: A Conceptual Review for Applied Psychologists. *Personnel Psychology*. 47, S. 537 – 560.

Salmon, W. (1998): Causality and Explanation. New York: Oxford University Press.

Scherer, K. und Zentner, M. (2001): Emotional Effects of Music: Production Rules. S. 361 – 392. In: Juslin, P. und Sloboda, J. (Hg.): Music and Emotion. Oxford: Oxford University Press.

Schermelleh-Engel, K., Moosbrugger, H. und Müller, H. (2003): Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures. *Methods of Psychological Research Online*. 8(2), S. 23 – 74.

Schnell, R.; Hill, P. und Esser, E. (2005): Methoden der empirischen Sozialforschung. München: Oldenbourg.

Schoenmakers, H. und Tulloch, J. (2004): From Audience Research to the Study of Theatrical Events: A Shift in Focus. S. 15 – 27. In: V. Cremona, P. Eversmann, H. van Maanen, W. Sauter und J. Tulloch (Hg.): Theatrical Evens – Borders, Dynamics, Frames. Amsterdam: Rodopi.

Schugk, M. (1996): Betriebswirtschaftliches Management öffentlicher Theater und Kulturorchester. Wiesbaden: Deutscher Universitätsverlag.

Schulten, M. (1990): Musikpräferenz und Musikpädagogik. Frankfurt a.M.: Lang.

Sievers, N. (2005): Publikum im Fokus. Begründungen einer nachfrageorientierten Kulturpolitik. In: Institut für Kulturpolitik der Kulturpolitischen Gesellschaft (Hg.): Jahrbuch für Kulturpolitik 2005. Essen: Klartext Verlag.

Sloboda, J. (1991): Music Structure and Emotional Response: Some Empirical Findings. *Psychology of Music*. 19, S. 110 – 120.

Sloboda, J. (1999): Musical Performance and Emotion: Issues and Development. S. 220 – 238. In: Yi, S. (Hg.): Music, Mind, and Science. Seoul: Seoul National University Press.

Sloboda, J. und O'Neill, S. (2001): Emotions in Everyday Listening to Music. S. 415 – 429. In: Juslin, P. und Sloboda, J. (Hg.): Music and Emotion. Oxford: Oxford University Press.

Sowa, J., Selden, S. und Sandfort, J. (2004): No Longer Unmeasurable? A Multidimensional Integrated Model of Nonprofit Organizational Effectiveness. *Nonprofit and Voluntary Sector Quarterly.* 33(4), S. 711 – 728.

Statistisches Bundesamt (2006): Kulturfinanzbericht 2006. Wiesbaden.

Stevens, S. (1946): On The Theory of Scales of Measurement. *Science*. 103, S. 677 – 680.

Thompson, S. (2006): Audience Responses to a Live Orchestral Concert. *Musicae Scientiae*. 10(2), S. 215 – 244.

Thompson, S. (2007): Determinants of Listeners' Enjoyment of a Performance. *Psychology of Music*. 35 (1), S. 20 – 36.

Thompson, S. und Williamon, A. (2003): Evaluating Evaluation: Musical Performance Assessment as a Research Tool. *Music Perception*. 21(1), S. 21 – 41.

Upmeyer, A. (1985): Soziale Urteilsbildung. Stuttgart: Kohlhammer.

Wapnick, J. und Ekholm, E. (1997): Expert Consensus in Solo Voice Performance Evaluation. *Journal of Voice*. 11, S. 429 – 436.

Weaver, J. (2000): Personality and Entertainment Preferences. S. 235 – 248. In: Zillmann, D. und Vorderer, P. (Hg.): Media Entertainment. The Psychology of its Appeal. London: Lawrence Erlbaum.

Wilson, G. (1985): The Psychology of the Performing Arts. London, Sydney: Croom Helm.

Winter, N. (1993): Music Performance Assessment: A Study of the Effects of Training and Experience on the Criteria used by Music Examiners. *International Journal of Music Education*. 22, S. 34 – 39.

Wirtz, M. und Caspar, F. (2002): Beurteilerübereinstimmung und Beurteilerreliabilität. Göttingen: Hogrefe.

Zillmann, D. (1991): Empathy: Affect from Bearing Witness to the Emotions of Others. S. 135 – 167. In: Bryant, J. und Zillmann, D. (Hg.): Responding to the Screen: Reception and Reaction Processes. Hillsdale, NJ: Erlbaum.

Zillmann, D. (1994): Mechanisms of Emotional Involvement with Drama. *Poetics*. 23, S. 33 – 51.