

**EVALUATION IN MUSIC THEATER: EMPIRICAL
RESULTS ON CONTENT AND STRUCTURE OF
THE AUDIENCE'S QUALITY JUDGMENT**

SABINE BOERNER

University of Constance, Germany

HANS NEUHOFF

Cologne University of Music, Germany

SABINE RENZ

VOLKER MOSER

University of Constance, Germany

ABSTRACT

This article investigated audience judgment of performance quality in opera. On the basis of Boerner's (2004) componential framework of performance quality in opera, a first version of the "questionnaire for the perception of performance quality in music theater" was developed. This questionnaire was validated in two pilot studies with 70 and 39 graduate students in Studies 1 and 2, respectively. In Study 3, the content and the structure of spectators' judgments of performance quality in opera were examined in a field study by interviewing 145 visitors to the opera *La forza del destino* (Verdi) in Cologne Opera House. Conclusions for further research on reception of opera performance are drawn.

Music theater, combining the presentation of music, language/plot and manifold visual stimuli, undoubtedly makes for one of the most complex and demanding forms of aesthetic experience, enjoying, among other reasons, high cultural status

from that (Dollase, Rösenberg, & Stollenwerk, 1986). While quality in opera is generally considered a dimension that is difficult to measure (Allmann, 1997; Hoegl, 1995; Lipp, 1994; Tobias, 2004), value judgments of quality of performance in opera are regularly being produced by visitors, critics, artists, and other professionals. However, to date, neither culture-political or management-focused research, nor music-psychological research, have provided a scientific tool which is able to grasp the performance quality of an opera company. Our project, therefore, conducts fundamental research in the field, aiming to answer the question, What is the subject of the evaluation of an opera performance? This question implies an analysis of the content and the structure of manifest evaluative judgments. Specifically, we investigated two research questions: 1) Which components of both musical (e.g., orchestra, soloists) and staging dimension (e.g., stage set, costumes) of an opera performance do spectators perceive? and 2) How are these different components weighted within spectators' overall judgments of performance quality in opera?

Research in music psychology has developed a range of instruments to measure the quality judgment of individual players of instruments (Boyle & Radocy, 1987; Burnsed, Hinkle, & King, 1985; Fiske, 1979; Wapnik & Ekholm, 1997), individual singers (Kleber, 2004), and university orchestras and choirs (Cooksey, 1982; Sagen, 1983). Pavis (1988) suggested criteria for the analysis of theater performances to be used in students' education. Sauter's "theater talks" (Sauter, Kalvik, & Isaksson, 1986) constitute a qualitative approach in reception research. Perkey (1976) developed a scale for the evaluation of acting quality. The only attempt to capture the quality judgment on an entire opera performance has been presented by Behr (1983). His study, however, does not focus on the quality judgment as such, but on the different effects (e.g., emotional, aesthetic) an opera performance has on the audience.

To the best of our knowledge, no attempt has been made so far to develop an instrument to investigate the content and the structure of audience's judgments of performance quality in an opera performance. The aim of this research was to help close this gap by conducting fundamental research in the evaluation of opera performances.

Since this is the first attempt to investigate the audience's quality judgment, we used a combination of qualitative and quantitative methods. In the first step, we developed a first version of a "questionnaire for the perception of performance quality in music theater." We content-analyzed professional reviews of opera-performances found in three nation-wide German newspapers in terms of the applied criteria. The results of this qualitative analysis were aligned with quality criteria stemming from Boerner's (2004) suggestion for a conceptualization of performance quality in opera. In the second step, this questionnaire was validated in two pilot studies with graduate students, presenting cuts of the operas *Aida*, *La Bohème*, and *The Mastersingers of Nuremberg*. Thirdly, the questionnaire was taken to the field (*La Forza del destino*) and first answers to the research

question were generated. In the following, we report on the development of the instrument and the empirical studies, drawing conclusions for further research on reception in opera.

DEVELOPING THE INSTRUMENT

In her componential framework of performance quality in opera, Boerner (2004) distinguishes between musical dimension (orchestra, choir, soloists) and staging dimension (acting quality, staging quality; see Table 1). The system can be refined by differentiating these components. On the musical dimension, for example, the component "orchestra" can be differentiated by instrument groups (strings, wind, percussion), for each group by individual instruments (e.g., violin, viola), and for each instrument by part (1st or 2nd violin). The choir as a component can be differentiated by voice (soprano, alto, tenor, and bass) and part (e.g., 1st or 2nd soprano). Similarly, soloists can be differentiated by voice, with the addition of mezzo-soprano and baritone. On the staging dimension, according to Fischer-Lichte (1983), the acting quality can be divided into the performer's activity (language, expression, gestures, movement) and his appearance (makeup and hairstyle). Staging quality includes costumes and stage set (spatial conception and stage space).

However, owing to the multi-dimensional nature of the stage work (Koebner, 1993) and the complexity of the differentiated dimensions, judging artistic quality in opera goes beyond the isolated quality of the components, like, for example, the sound of the orchestra or the soloists' articulation. Equally important for the

Table 1. Componential Framework of Performance Quality in Opera (Boerner, 2004)

First-order fit (fit 1) Fit between musical dimension and staging dimension				
<i>Musical dimension</i>			<i>Staging dimension</i>	
Second-order fit (fit 2) Fit within the musical dimension			Second-order fit (fit 2) Fit within the staging dimension	
Orchestra quality (instruments, parts)	Quality of the choir (parts)	Quality of the soloists' parts	Acting quality (behavior, appearance)	Staging quality (costumes, spatial concept, stage space)
Third-order fit (fit 3)	Third-order fit (fit 3)	Third-order fit (fit 3)	Third-order fit (fit 3)	Third-order fit (fit 3)

performance quality is the congruity of the components (Adorno, 1968, p. 12; Dahlhaus, 1973, p. 11). Congruity is specified in this model on three levels and referred to as "fit" (see MacInnis & Park, 1991; North & Hargreaves, 1997).

Fit is firstly important between the musical and staging dimension (*first-order fit* or "*fit1*"). For example, the tonal image produced in the musical dimension should match the atmosphere conveyed in the staging dimension. Decisive for the first-order fit of a performance is the guiding idea which the specific artistic conception takes as the basis for a production. Secondly, fit must be achieved within each dimension (*second-order fit* or "*fit2*"). Within the musical dimension, the individual components (orchestral quality, chorus quality, quality of solo singing) should harmonize with respect to sound (e.g., intonation, articulation, dynamics). The same is true for the components of the staging dimension: acting quality and costumes, for example, have to harmonize with respect to the atmosphere. Thirdly, third-order fit must be sought within the given components of a dimension, i.e., within the orchestra or within staging quality (*third-order fit* or "*fit3*"). The entry of the individual orchestral musicians must be synchronized, for example. Similarly, costumes, spatial concept, and stage space must be coordinated with the overall staging conception.

In our study, we used this componential framework to identify aspects that spectators might perceive in an opera performance and consider in their judgments of performance quality. Although this framework has been discussed with several professionals in the field (e.g., musicians, conductors, singers, and critics), we did not know which of the components in this concept spectators *actually use* for their evaluations of performance quality in opera. Therefore, we analyzed expert reviews of opera performances in order to identify the criteria professional critics use in their judgments (Boerner & Moser, 2006; Moser, 2005). We selected three nation wide German newspapers with a renowned feature section (*Die Zeit*, *Frankfurter Allgemeine Zeitung*, *Süddeutsche Zeitung*). We content-analyzed all reviews of opera performances (June to August 2004) in terms of the criteria critics applied. This analysis included 82 reviews by 43 different critics, referring to 65 performances in 44 German opera companies.

We identified 27 different "categories" (e.g., costumes, orchestra sound, and soloist) that we could, for the most part, integrate into the componential framework. Some additional aspects were found, like faithfulness to the opus, room for the spectator's fantasy, and the originality of the orchestral performance, that have been integrated in the questionnaire when formulating the items. Almost all of the aspects from the componential framework have been found in these expert reviews. Only the choir and the "third-order fit" ("fit3") within the orchestra, within the choir, and within the soloists are hardly ever mentioned in the reviews.

To conclude, Boerner's (2004) componential framework of performance quality seems to reflect the framework used by professional theater critics for their evaluation of performance quality in music theater quite completely. Although we do not know if this "expert framework" is valid for the "typical" or "average"

audience in music theater, we used this framework (with only minor modifications) for developing a first version of the questionnaire.

The construct "performance quality in opera" can thus be conceptualized as multi-dimensional, including the following aspects (see Table 1): orchestra (including third-order fit), choir (including musical performance, staging performance, and third-order fit), soloists (including musical performance, staging performance, and third-order fit), staging (including third-order fit), fit within the musical dimension (second-order fit), fit within the staging dimension (second-order fit), and fit between staging dimension and musical dimension (first-order fit). Since the conductor is expected to have a considerable impact on the overall performance quality in opera (Boerner, 2002; Boerner & von Streit, 2005), we included the construct "conductor" in the questionnaire. Moreover, a scale for the construct "overall performance" was developed and added. Several demographic variables (age, sex, general experience in opera, knowledge of the opera presented) that have proved to be relevant for persons' musical preferences (Behr, 1983; Neuhoff, 2001) were also added (see Table 2).

In total, the first version of the questionnaire for the perception of performance quality in music theater (see Table 2) included 1) the individual aspects (orchestra, choir, soloists, scenery, conductor), 2) the interplay within (first-order fit, second-order fit, third-order fit) and between these aspects (e.g., fit between orchestra and soloists), 3) the overall performance, and 4) demographic variables. Examples for the operationalization of these constructs are given in Table 3.

VALIDATION OF THE INSTRUMENT: PILOT STUDIES

Participants

In Study 1 (Moser, 2005), our sample consisted of 70 students at the University of Constance, who were either members of the University orchestra and choir ($n = 40$) or participants of a seminar on theater management in 2005 ($n = 30$). Due to this selection, only a small number (18%) of the participants rated themselves as "non-experts"; nearly half of our sample (49%) had "some experience"; for nearly a quarter (24%) music theater was a hobby, and 7% answered to be "almost expert" or "expert" in music theater. More than one third (37%) of the participants were familiar with the operas we presented in the study. The majority (60%) of our respondents was female; the average age was 25 years. Study 2 (Renz, 2006) was conducted with a sample of 39 students at the University of Constance using the same setting and stimuli as in Study 1. Twenty-three percent of the participants rated themselves as "non-experts"; almost two-thirds (64%) had "some experience," and for 10% music theater was a hobby. Nearly half (44%) of the participants were familiar with the operas we presented in the study. More than half (59%) of our respondents were female, and the average age of the sample was 24.

Table 2. The Questionnaire for the Evaluation of Performance Quality in an Opera Company—Overview

A. Overall performance	
B. Individual aspects	
Musical performance of the ensemble	Staging/acting performance of the ensemble
orchestra (musical performance), soloists (individual performance of the protagonist), choir (musical performance), conductor	soloists (acting performance of the protagonist), choir (acting performance); staging/scenery (stage, costumes, decoration, properties, lighting)
C. Fit within and between the aspects	
fit within the orchestra (fit3-orchestra); fit within the choir (fit3-choir); fit between the soloists (fit3-soloists); fit between orchestra, soloists, and choir (fit2-music); fit within the staging dimension (fit2-stage), fit between music and stage (fit1)	
D. Demographical variables	

Note: According to existing research on the evaluation of musical performance (Boyle & Radocy, 1987; Cooksey, 1982), each of these aspects was measured using 5-point Likert-scales, consisting of 3 to 17 items each (see Table 1).

Stimuli and Procedure

In order to check for possible differences in the subjects' perceptions that can be attributed to the opera presented (Behr, 1983), we used performances of different operas and different opera companies as stimuli. The participants were shown 20-minute videos of three opera performances: *Aida* (Metropolitan Opera), *La Bohème* (San Francisco Opera), and *The Mastersingers of Nuremberg* (Zurich Opera). The videos each consisted of one or more subsequent cuts of the very opera (e.g., *La Bohème*: the complete 2nd act). Our criteria for the selection of the cuts was to include, if possible, the conductor (e.g., in a part of the overture), the choir, soloists' solo parts and ensemble parts (e.g., duet, tercet, soloist, and choir) in the video. In order to avoid an average judgment on the soloists' quality, our instruction was to concentrate exclusively on one soloist in every opera (*Aida*, Rodolfo, Hans Sachs).

The participants were provided with a short description of the plot selected for the presentation, and with basic information about the opera company and the cast. The questionnaires were filled out immediately after each video presentation; the whole experiment took about 2½ hours.

Table 3. The Questionnaire for the Perception of Performance Quality in Music Theater—Constructs and Exemplary Items

Constructs (number of items)	Exemplary items
OP (17)	I think the performance was a total success.
MO (9)	The orchestra elaborated differences in volume well.
MS (13)	The soloist's voice was little beautiful in terms of sound. ^R
MC (8)	The choir accentuated differences in volume very well.
C (18)	The conductor motivated the orchestra to a musical peak performance.
SS (11)	The soloist embodied the figure's characteristics very convincingly.
SC (5)	From its appearance, one could fully believe the choir's role.
S (12)	Stage setting and decoration distracted the plot with too many details. ^R
Fit3O (8)	The different instrumental groups in the orchestra (strings, blowers, percussion) coordinated well regarding the volume.
Fit3C (5)	The sound of the choir seemed balanced, so that single person's voices were never clearly identifiable.
Fit3S (7)	Concerning their charisma, the soloists did not fit each other well.
Fit2M (9)	Orchestra and choir fitted well regarding the tempo. The orchestra was neither slower nor faster than the choir.
Fit2S (12)	Costumes and stage setting matched in such a way that a "harmonious" atmosphere was developed.
Fit1 (8)	Altogether in this performance the "music" (orchestra, choir, and soloists) matched the "scene" (stage settings, decoration, props, costumes, lighting) well.

Note: OP = Overall performance, MO = Orchestra-musical performance, MS = Soloists-musical performance, MC = Choir-musical performance, C = Conductor, SS = Soloists-staging performance, SC = Choir-staging performance, S = Scenery-general stage setting, fit3O = Interplay within the orchestra, fit3C = Interplay within the choir, fit3S = Interplay between the soloists, fit2M = Interplay between orchestra, soloists and choir, fit2S = Interplay between the staging aspects, fit1 = Interplay between musical dimension and staging dimension; ^R = polarity reversed.

Results of Study 1

In the first step of the analyses, the reliabilities of all scales in the questionnaire proved to be acceptable, Cronbach's alphas being all well above .70 for all three operas with the exception of "soloists – musical performance (MS)" (see Table 4).

The most appropriate method to validate the multi-dimensional construct "performance quality in opera" would be a confirmatory factor analysis. For this procedure, however, the samples of our pilot studies are too small (Kline, 2004). We therefore conducted explorative factor analyses in order to validate the construct. Due to the hierarchical nature of our underlying quality framework (see Table 1), we did not expect all of the quality components to be confirmed in one single analysis. Instead, we decided to analyze the three levels of the framework separately. The first-order fit was confirmed to form one single factor in all three operas (explained variance varying from 52% in *La Bohème* to 59% in *Aida*). On the second level, the two second-order fits, namely fit2-music and fit2-stage, were confirmed to form two different factors in all three operas (see Table 5 for *Aida* as an example). On the third level, however, for none of the three operas the complete set of the constructs that have been identified as reliable in the first step could be confirmed in factor analyses.

Results of Study 2

Based on Study 1, the questionnaire was improved (items with factor loadings $< .30$ and with cross-loadings were removed, some items were reworded). Additionally, the response scale for all items was enlarged by the alternatives "I did not pay attention to" and "I am not able to judge." With these options, we expected to capture the audience's "real" judgment more precisely than in Study 1, avoiding arbitrary answers in case of respondents' uncertainty.

Descriptive analyses of the data revealed that the two additional response options were used to a considerable degree (Boerner, Neuhoff, Moser, & Renz, 2006; Boerner, Neuhoff, & Renz, 2006). Therefore, we excluded items which at least 30% of our respondents did not notice or were not able to judge from further analyses. "Missing values" in the remaining items were imputed using maximum likelihood approximations (Roth, 1994). Nevertheless, adding the two alternative response options resulted in a complete loss of the two scales "conductor" and "third-order fit – orchestra" due to insufficient reliability. All other constructs of the questionnaire showed acceptable reliabilities (see Table 4).

To validate the multi-dimensional construct "performance quality in opera," we proceeded as in Study 1. The first-order fit was confirmed to form one single factor in all three operas (explained variance varying from 49% for the opera *Aida* to 68% for the opera *La Bohème*). On the second level, the two second-order fits, namely fit2-music and fit2-stage, were confirmed to form two different factors in all three operas (see Table 6 for the example *Aida*). On the third level, only for the opera *Aida* could all four constructs, orchestra, soloists, choir, and staging/scenery be identified as separate factors (see Table 7 for the example *Aida*). In the other

Table 4. Reliabilities of Scales (Cronbach's Alpha) in the Pilot Studies

Scale	<i>Aida</i>		<i>La Bohème</i>		<i>Mastersingers</i>	
	1st study	2nd study	1st study	2nd study	1st study	2nd study
MO	.82 (8/9)	.85 (3/3/9)	.83 (8/9)	.71 (3/3/9)	.83 (5/9)	.74 (4/4/9)
MS	.87 (12/13)	.65 (3/3/9)	.81 (9/13)	.80 (5/7/9)	.85 (9/13)	.89 (8/8/9)
MC	.84 (8/8)	— (1/1/5)	.89 (8/8)	.87 (4/4/5)	.91 (6/8)	.63 (3/3/5)
C	.89 (11/18)	— (-/-/10)	—	—	.80 (14/18)	— (-/-/10)
SS	.85 (8/11)	.69 (6/7/9)	.82 (9/11)	.81 (8/9/9)	.84 (7/11)	.88 (7/8/9)
SC	.81 (4/5)	.85 (4/5/6)	.79 (4/5)	.80 (4/6/6)	.90 (4/5)	.85 (5/6/6)
S	++	.85 (3/3/3)	++	— (2/2/3)	++	— (2/2/3)
Fit3O	.83 (7/8)	— (-/-/9)	.88 (7/8)	— (-/-/9)	.85 (4/8)	— (2/2/9)
Fit3C	.73 (4/5)	— (1/1/3)	.80 (5/5)	— (2/2/3)	.86 (3/5)	.85 (3/3/3)
Fit3S	.80 (5/7)	.71 (4/5/6)	.81 (4/7)	.79 (6/6/6)	.74 (4/7)	.73 (3/4/6)
Fit2M	.86 (8/9)	.79 (6/6/8)	.86 (7/9)	.84 (6/6/8)	.90 (6/9)	.83 (6/6/8)
Fit2S	.88 (10/12)	.86 (8/9/9)	.88 (12/12)	.84 (8/9/9)	.89 (9/12)	.95 (8/9/9)
Fit1	.89 (8/8)	.78 (6/6/6)	.84 (7/8)	.89 (6/6/6)	.88 (5/8)	.83 (6/6/6)
OP	++	.92	++	.94	++	.94
		(15/15/16)		(16/16/16)		(16/16/16)

Note: First study: $n = 70$; second study: $n = 39$. Values = Cronbach's alpha (number of items included in the final scale / Study 2: number of items remaining if all items that at least 30% of our sample did not notice or were not able to judge were excluded / number of items in the original scale); — the conductor was not shown in the video; ++ no scales due to modifications of the constructs. Legend: MO = Orchestra – musical performance, MS = Soloists – musical performance, MC = Choir – musical performance, C = Conductor, SS = Soloists – staging performance, SC = Choir – staging performance, S = Scenery – general stage setting, fit3O = Interplay within the orchestra, fit3C = Interplay within the choir, fit3S = Interplay between the soloists, fit2M = Interplay between orchestra, soloists and choir, fit2S = Interplay between the staging aspects, fit1 = Interplay between musical dimension and staging dimension, OP = Overall performance.

Table 5. Factor Analysis—Second Level—*Aida* (Study 1)

Item	Label	Factor loading	
		fit2-music	fit2-stage
fit2_3	Adjustment volume	.84	.28
rfit2_6	"Drowning out" choir	.80	.12
fit2_8	Balance – orchestra/soloists	.79	.21
rfit2_9	"Overflow" soloists	.79	.01
fit2_2	Voice – soloists	.66	.04
fit2S_1	Stage setting – plot	-.01	.87
fit2S_2	Scenery set/requisites – plot	.20	.76
rSA7	Distraction stage setting/scenery set	.13	.60
Explained variance		38%	23%

Note: $n = 70$. Principal component analysis (varimax rotation). Legend: fit2-music = Interplay between orchestra, soloists, and choir, fit2-stage = Interplay between the staging aspects. Values in bold indicate main factor loadings.

Table 6. Factor Analysis—Second Level—*Aida* (Study 2)

Item	Label	Factor loading	
		fit2-music	fit2-stage
fit2_3	Adjustment volume	.91	-.02
ufit2_7	"Harmonious sound characteristics"	.90	-.16
fit2_2	Voice – soloists	.69	.04
fit2S_1	Stage setting – plot	-.06	.92
fit2S_2	Scenery set/requisites – plot	-.01	.91
Explained variance		42%	34%

Note: $n = 39$. Principal component analysis (varimax rotation). Legend: fit2-music = Interplay between orchestra, soloists, and choir, fit2-stage = Interplay between the staging aspects. Values in bold indicate main factor loadings.

Table 7. Factor Analysis—Third Level—*Aida* (Study 2)

Item	Label	Factor loading			
		Scenery	Choir – staging	Soloists	Orchestra
fit3A2	Scenery – aesthetics	.96	.05	-.04	.13
SA13	Stage setting	.93	.08	-.09	.07
SA14	Stage setting technique	.83	-.14	-.08	.27
SA15	Costumes	.82	.18	-.21	-.04
SS1	Soloist – character	.67	.28	.28	.04
uSC4	Choir – motion	.07	.95	.02	.00
SC2	Choir – part	.05	.87	.06	.12
uSC1	Choir – appearance	.20	.75	.05	.02
SC3	Choir – aesthetics	-.01	.70	-.26	.18
rfit3S3	Soloists – “against each other”	-.09	.05	.88	-.03
ufit3S5	Soloists – “harmonious”	-.01	-.04	.87	-.04
fit3S1	Soloists – adjustment	-.06	-.01	.83	.09
MO2	Orchestra – volume	.06	.12	-.00	.91
MO10	Orchestra – “lively”	.11	.03	-.08	.90
MO5	Orchestra – tension	.13	.13	.11	.78
Explained variance		25%	19%	16%	16%

Note: $n = 39$. Principal component analysis (varimax rotation). Values in bold indicate main factor loadings.

two operas, however, only three constructs, namely, soloists, choir, and staging/scenery, could be identified as separate factors.

DISCUSSION

Our questionnaire seems to be valid in principle since most of the constructs can be identified in factor analyses. However, the constructs “conductor” and “fit3-orchestra” were not validated so far. An explanation for this result may be that the audience at an opera performance generally feels more competent to assess the staging dimension than particular aspects of the music dimension. As was assumed in Boerner’s (2004) componential framework, the construct “performance quality in opera” seems to be hierarchical in that three different

levels were identified in factor analyses. However, on the third level of the model, the full set of constructs was confirmed only in Study 2 and only for the opera *Aida*. In part, the reason for these shortcomings of the pilot studies may be their experimental character. Compared to a “real world” opera performance, there were some considerable deviations in our setting.

First, the live event was “replaced” by a video presentation, creating a different atmosphere. There was no interaction between the audience and the artists, and each of the operas presented was reduced to a small (and sometimes interrupted) cut of 20 minutes. Consuming only selected parts from three different operas may have created an artificial atmosphere. Since atmospheric and emotional aspects are assumed to have an impact on the individual judgment of artistic quality (Sauter et al., 1986; Williamon & Davidson, 2002), the particular setting of our studies may have caused some bias in the results. Second, our sample was certainly not “representative” of the audience for opera performances, especially concerning age and experience in opera. The audiences at operas are generally much older and have more experience in opera than the participants of our studies (Behr, 1983; Neuhoff, 2001) had. Third, as to the quality of our data, there were considerable “missing values” in Study 2, due to the extended response scale. The participants in our studies had some difficulties answering the items since they did not notice the aspects or were not able to judge the aspects. This may be one of the reasons why the construct “conductor”—undoubtedly relevant for the performance quality in opera—could not be identified in the analyses.

FIELD STUDY: FIRST EMPIRICAL RESULTS

Participants and Setting

We asked 145 spectators at a performance of *La forza del destino* in Cologne Opera House in May 2006 to complete our questionnaire. As in the pilot studies, our instruction was to concentrate exclusively on one soloist (Leonora). Similar to Behr’s (1983) procedure, the questionnaires accompanied by stamped envelopes were distributed randomly during the intermission and after the performance. The participants were predominately female (52%), the average age being 56 years. Five percent of the participants rated themselves as complete layman, 25% had some experience in opera, 50% considered opera to be their hobby, 10% rated themselves to be semi-professionals, and 6% rated themselves to be experts in opera.

Results

Descriptive analyses revealed that the respondents of this field study chose the new response options (“I did not pay attention to” and “I am not able to judge”) far less often than did the respondents in pilot Study 2 (Neuhoff, Boerner, & Renz, 2006). This may be due to the fact that the spectators were much older

($m = 56$ years) than the participants in the pilot studies ($m = 25$ and $m = 24$ years, respectively), and had more general experience in opera and better knowledge of the presented opera. This finding allows excluding less items from further analysis than in the pilot studies; we decided to exclude items which at least 10% of our participants did not notice or were not able to judge (Roth, 1994). The reliabilities of the scales were much better than in the pilot studies (see Table 8). However, similar to Study 2, the constructs "conductor" and "fit3-orchestra" did not reach acceptable reliabilities. Consequently, when answering the research questions these two components were not considered. "Fit3-soloists" consists of two items only (Boerner et al., 2006; Renz, 2006).

To answer the first research question, which components of both musical and staging dimension of an opera performance spectators perceive, we conducted separate factor analyses for each of the levels of the model. As in the pilot studies, fit1, fit2-music and fit2-stage (explained variances are 70%, 26%, and 45%, respectively) could be identified as separate factors. On the third level, the following six factors could be identified (see Table 9): soloists-music, choir-staging, orchestra, soloists-staging, choir-fit3, and soloists-fit3. However, the last factor consists of only 2 items.

To determine how the different components are weighted within spectators' overall judgments of performance quality (question 2)—the second research question—we conducted a path analysis (AMOS 5.0) predicting the spectators' overall judgments of the performance of *La forza del destino* using the components of the construct "performance quality in opera."

The path analysis confirmed that a hierarchical model ($\chi^2 = 17.82$; $p = .121$; CMIN/DF = 1.485; NFI = .966; RMSEA = .058; $p_{close} = .362$; see Figure 1)¹ fits better to the data than an alternative general factor model ($\chi^2 = 418.02$; $p = .001$; CMIN/DF = 27.868; NFI = .206; RMSEA = .432; $p_{close} = .001$). All three levels of the componential framework could be reconstructed, explaining 61% of the variance in the spectators' overall judgments. However, on the third level of the model, not all constructs that had been identified in the factor analysis were reconstructed: The constructs soloists-music, orchestra, and soloists-staging are missing. (Since we assumed the overall performance may not exclusively be affected by fit1, as suggested in the hierarchical framework, but by any of the other components of the construct "performance quality in opera," we alternatively allowed additional *direct* paths from all components to the "overall performance." However, none of these direct paths turned out to be significant.) As shown in the path model, fit1 has the highest weight on the overall performance (total effect = .78). The weights of the other quality components on the overall performance are shown in the total effects in Table 10.

¹ χ^2 (CMIN) = Chi-Square-Value, p = probability level, df = degrees of freedom, NFI = Normed Fit Index, RMSEA = Root Mean Square Error of Approximation; p_{close} = probability level for a higher value as the empirical RMSEA (close fit) (Schermelleh-Engel & Moosbrugger, 2003).

Table 8. Reliabilities (Cronbach's Alpha) of All Scales—
La forza del destino (Field Study)

Scale	Cronbach's alpha	Scale	Cronbach's alpha
MO	.88 (3/3/9)	fit3O	— (-/-/9)
MS	.93 (7/8/9)	fit3C	.87 (3/3/3)
MC	.87 (3/3/5)	fit3S	.82 (4/5/6)
C	— (-/-/10)	fit2S	.96 (7/8/9)
SS	.89 (8/8/9)	fit2M	.87 (6/6/8)
SC	.91 (5/6/6)	fit1	.89 (5/5/8)
S	— (2/2/3)	OP	.95 (16/16/16)

Note: $n = 145$. Values = Cronbach's alpha (number of items included in the final scale / number of items remaining if all items that at least 10% of our sample did not notice or were not able to judge were excluded / number of items in the original scale). Legend: MO = Orchestra – musical performance, MS = Soloists – musical performance, MC = Choir – musical performance, C = Conductor, SS = Soloists – staging performance, SC = Choir – staging performance, S = Scenery – general stage setting, fit3O = Interplay within the orchestra, fit3C = Interplay within the choir, fit3S = Interplay between the soloists, fit2M = Interplay between orchestra, soloists, and choir, fit2S = Interplay between the staging aspects, fit1 = Interplay between musical dimension and staging dimension, OP = Overall performance.

To sum up, for the overall judgment on the performance of *La forza del destino*, the most important components were fit1, fit2-stage and fit3-choir, whereas fit2-music, choir-staging and fit3-soloists were of minor importance.

DISCUSSION

First of all, the questionnaire seems in principle to stand the test in the field, for most of the scales improved in reliability (Cronbach's alpha; see Table 8) compared to the pilot studies (see Table 4). At the same time, compared to the pilot studies (see Table 7), the factor analysis on the third level of the quality model produces more individual components in the field study (see Table 10). However, similar to the pilot studies, the scales "conductor," "fit3-orchestra," and "staging/scenery" were not confirmed in the field study due to insufficient alpha coefficients and the scale "fit3-soloists" consists only of two items. Therefore, the questionnaire clearly needs further improvement.

Moreover, the path model includes only the constructs "fit3-soloists," "fit3-choir," and "choir-staging" on the third level. One could argue that a judgment of performance quality in opera which excludes orchestra, conductor and soloists-music staging, is doubtful in itself. However, this result has to

Table 9. Factor Analysis—Third Level—*La forza del destino*
(Field Study)

Item	Factor loading					
	Soloists – music	Choir – staging	Orchestra	Soloists – staging	Choir – fit3	Soloists – fit3
uMS7	.87	.15	.12	.11	.14	-.00
MS6	.83	.07	.18	.14	.11	.12
MS5	.82	.07	.19	.26	.20	.13
MS10	.81	.03	.17	.27	.25	.07
MS1	.77	.01	.20	.15	-.00	.21
uSC1	-.00	.92	.05	.06	.08	.02
SC2	.06	.86	.17	.19	.16	.08
SC3	.00	.86	.09	.09	.16	.07
uSC4	.24	.80	.10	.12	.09	.08
MO5	.22	.10	.87	.15	.15	.19
MO10	.26	.09	.84	.13	.16	.11
uMO3	.23	.23	.80	.10	.24	-.03
SS5	.24	.21	.01	.88	.09	.13
SS4	.26	.24	.13	.85	.06	.03
SS1	.24	.03	.25	.83	.16	.08
ufit3C2	.13	.19	.26	.17	.85	.05
fit3C3	.26	.13	.00	.03	.80	.25
fit3C1	.15	.20	.37	.14	.80	.02
urfit3S6	.11	.13	.06	.05	.05	.91
fit3S1	.29	.07	.22	.19	.27	.70
Explained variance	20%	16%	13%	13%	12%	8%

Note: $n = 145$. Principal component analysis (varimax rotation). Values in bold indicate main factor loadings.

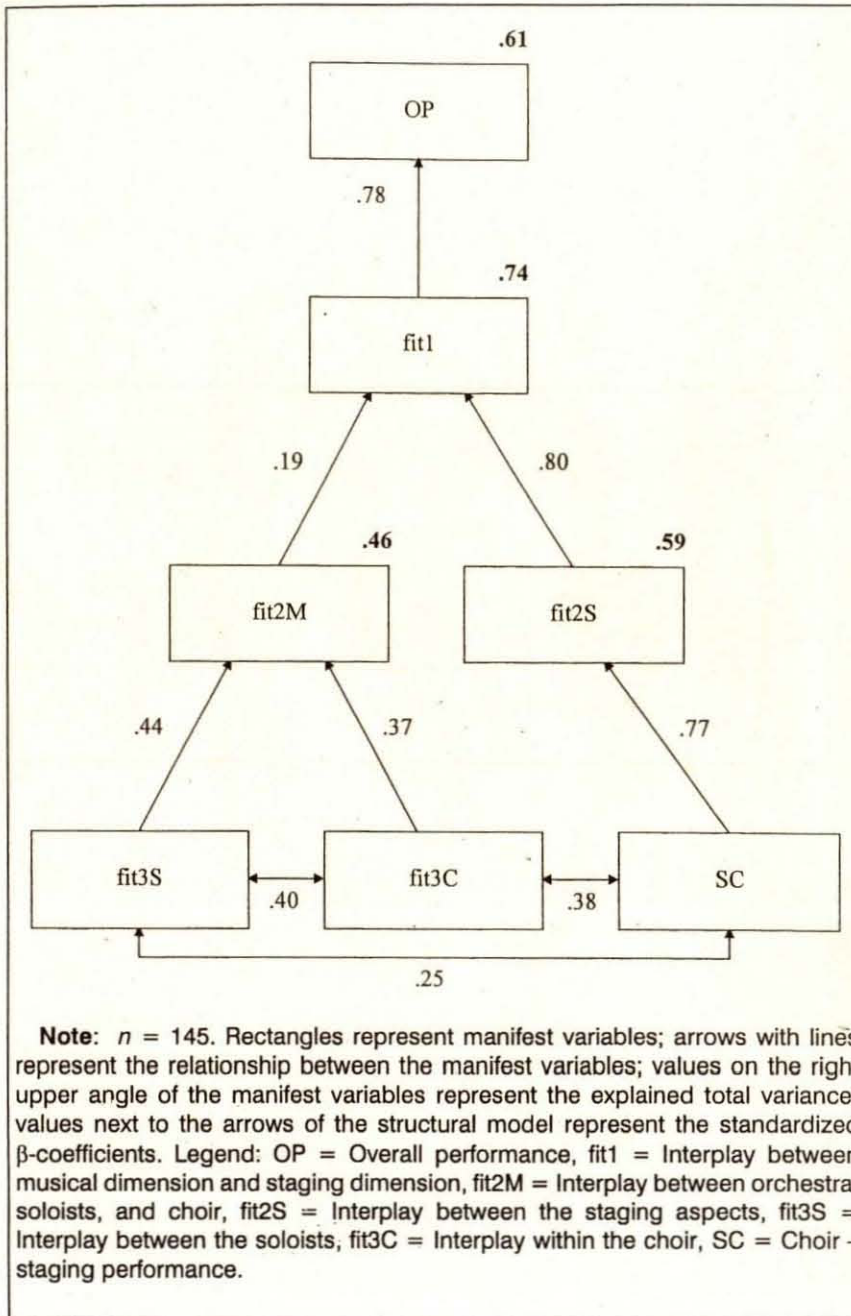


Figure 1. Path model for the opera *La forza del destino*.

Table 10. Standardized Total Effects of Overall Performance
 "La forza del destino" (Field Study)

	fit3S	fit3C	SC	fit2S	fit2M	fit1
OP	.06	.48	.05	.63	.15	.78

Note: $n = 145$. Principal component analysis (varimax rotation). Total effects are the sum of direct and indirect effects. Legend: OP = Overall performance, fit3S = Interplay between the soloists, fit3C = Interplay within the choir, SC = Choir – staging performance, fit2S = Interplay between the staging aspects, fit2M = Interplay between orchestra, soloists and choir, fit1 = Interplay between musical dimension and staging dimension.

be interpreted carefully, since the hierarchical structure of the construct "performance quality in opera" (see Table 1) necessarily implies redundancy: Asking for "fit1," "fit2-music," and "fit2-stage" includes *per definitionem* the performance of conductor, orchestra, soloists, and staging/scenery. Since all components of the multidimensional construct "performance quality in opera" are assumed to be interdependent, it is necessarily difficult to identify individual aspects. For example, it is difficult to operationalize the construct "conductor" without mentioning the orchestra or the soloists in the items. Although orchestra and soloists-music are missing, in the path model 46% of the variance in fit2-music can be explained (see Figure 1).

Apart from the above limitations of this study, the research questions can be answered as follows: 1) In an opera performance, spectators perceive individual components of the musical dimension (orchestra, soloists, and choir) and of the staging dimension (soloists-staging, choir-staging). Moreover, they perceive the congruency between musical and staging dimensions ("fit1"), the fit within the musical dimension ("fit2-music"), the fit within the staging dimension ("fit2-stage"), the fit within the choir ("fit3-choir") and within the soloists ("fit3-soloists"), whereas the fit within the orchestra ("fit3-orchestra") was not confirmed as a separate construct in our studies. One reason may be that the fit within the orchestra, being a very complex quality aspect, is difficult to judge in an opera performance by non-professionals. 2) For the overall judgment of an opera performance, the most important aspect is fit1, followed by fit2-stage, and fit3-choir. For the opera under study, the staging dimension (standardized β -coefficient = .80; see Figure 1) seems to be far more relevant for "fit1" than the musical dimension (standardized β -coefficient = .19; see Figure 1). One possible interpretation of this result is the following. In the eyes of the spectators, the musical dimension of an opera performance is perceived as "given," as more or less "constant" or guaranteed. Spectators probably consider the musical dimension as such to be closer to the original opera, whereas the staging dimension is perceived as an interpretation that may go beyond the original. Hence, in their judgment of the fit between musical dimension and staging dimension (fit1),

spectators focus primarily on the question of whether the staging dimension fits the music than the other way round.

SUMMARY AND FURTHER RESEARCH

The "questionnaire for the perception of performance quality in music theater" that was developed in the presented studies seems to be a valid instrument in principle to measure an audiences' judgment of the quality in opera. On the basis of two pilot studies and one field study, the answers to the research questions are as follows: 1) In their judgment on the quality of an opera performance, spectators perceive both individual aspects (e.g., the soloists) and congruency aspects (e.g., the fit between staging dimension and music dimension); and 2) As to the weighting of these aspects in spectators' overall judgments, the most important aspect seems to be fit1, followed by fit2-stage and fit3-choir. For the judgment on fit1, the staging dimension seems to be much more relevant than the musical dimension.

Further studies will have to improve the questionnaire, especially in order to properly measure the constructs "conductor" and "orchestra." In order to avoid excluding items from the analysis and computation procedures, the final version of the questionnaire should contain only aspects that spectators consider in their judgments and that they feel competent to assess. With the validity of the questionnaire improved, the two research questions may be answered systematically, applying the questionnaire to different performances of different operas. Since the operas included in our studies so far all belong to the mainstream, another interesting point is the question if characteristics of the piece or the production affect spectators' judgment of performance quality.

Moreover, further research questions may be answered. For example, it is interesting how different judges assess the quality of an opera performance. This question addresses the agreement in subjective judgments of performance quality (Lesle, 1981). One could analyze if and in which quality components different groups of judges (artists, critics, and audience/laities) agree in their judgments. Furthermore, regarding the differentiation of components and the weighting of these components in the quality judgment, what differences exist between experts and non-experts? Since experts have more experience and competence in opera, one could hypothesize that they perceive and consider more individual quality components in their judgment than do non-experts. Another important question is how much agreement there is in the quality judgment within the different groups; is there more agreement in experts' judgments than in non-experts' judgments of an opera performance?

REFERENCES

- Adorno, T. (1968). Reflexionen über Musikkritik [Reflexions on music review]. In H. Kaufmann (Ed.), *Symposium für Musikkritik* [Symposium on music review] (pp. 7-21). Graz: Institut für Wertungsforschung.

- Allmann, U. (1997). *Innovatives Theatermanagement. Eine Fallstudie* [Innovative theater management. A case study]. Wiesbaden: Deutscher Universitätsverlag.
- Behr, M. (1983). *Musiktheater—Faszination, Wirkung, Funktion* [Music theater—Fascination, impact, function]. Wilhelmshaven: Heinrichshofen.
- Boerner, S. (2002). *Führungsverhalten und Führungserfolg—Beitrag zu einer Theorie der Führung am Beispiel des Musiktheaters* [Leadership behavior and leadership success—Contribution to leadership theory taking the example of music theater]. Wiesbaden: Gabler.
- Boerner, S. (2004). Artistic quality in an opera company—Towards the development of a concept. *Nonprofit Management and Leadership*, 4(4), 425-436.
- Boerner, S., & von Streit, C. (2005). Transformational leadership and group climate—Empirical results from German symphony orchestras. *Journal of Leadership and Organizational Studies*, 12(3), 31-41.
- Boerner, S., & Moser, V. (2006). *Evaluating performance quality in an opera company the validation of an instrument and first empirical findings*. Paper presented at the Fourth International Conference on Cultural Policy Research, Vienna.
- Boerner, S., Neuhoff, H., Moser, V., & Renz, S. (2006). *Quality in the performing arts: Evaluation of performance in music theatre*. Paper presented at the Fifth International Conference on theory and practice in performance measurement, London.
- Boerner, S., Neuhoff, H., & Renz, S. (2006). *Evaluation im Musiktheater* [Evaluation in music theater]. Jahrestagung der Fachgruppe Systematische Musikwissenschaft der Deutschen Gesellschaft für Musikpsychologie [Annual meeting of the Fachgruppe Systematische Musikwissenschaft the German Society for Music Psychology], Kassel.
- Boyle, J. D., & Radocy, R. E. (1987). *Measurement and evaluation of musical experience*. New York: Schirmer Books.
- Burnsed, V., Hinkle, D., & King, St. (1985). Performance evaluations reliability at selected concert festivals. *Journal of Band Research*, 21(1), 22-29.
- Cooksey, J. N. (1982). Developing an objective approach to evaluating music performance. In R. Colwell (Ed.), *Symposium in music education* (pp. 197-229). Illinois: University of Illinois.
- Dahlhaus, C. (1973). Über musikalische Werturteile [On value judgments of music]. In W. Krütfeld (Ed.), *Wertproblematik der Musikdidaktik* [Value judgments in music didactics] (pp. 9-18). Ratingen: Henn.
- Dollase, R., Rüsenberg, M., & Stollenwerk, H. J. (1986). *Demoskopie im Konzertsaal* [Opinion poll in the concert hall]. Mainz: Schott.
- Fiske, H. E. (1979). Musical performance evaluation ability: Toward a model of specificity. In E. Asmus (Ed.), *Psychology and acoustics of music* (pp. 147-169). Lawrence, KS: University of Kansas Press.
- Fischer-Lichte, E. (1983). *Semiotik des Theaters* [Semiotics in theater] (Band 1: Das System der theatralischen Zeichen [The system of theatrical signs]). Tübingen: Narr.
- Hoegl, C. (1995). *Ökonomie der Oper. Grundlagen für das Musiktheater-Management* [Economics of opera. Fundamentals for music theater management]. Bonn: ARCult Media.
- Kleber, B. (2004). *Evaluation von klassischem westlichem Gesang* [Evaluation of classical western singing]. Unpublished master's thesis, University of Constance, Constance, Germany.

- Kline, R. B. (2004). *Principles and practice of structural equation modelling*. New York: Guilford Press.
- Koebner, T. (1993). *Handlungen mit Musik. Die Oper als Zeitspiegel, Leidenschaftsdrama, Gesamtkunstwerk* [Plot with music. Opera as reflection, drama, Gesamtkunstwerk]. Anif/Salzburg: Müller-Speiser.
- Lesle, L. (1981). *Notfall Musikkritik* [Music review as case of need]. Wiesbaden: Breitkopf & Härtel.
- Lipp, W. (1994). *Drama Kultur* [Culture as a drama]. Berlin: Duncker & Humblot.
- MacInnis, D. J., & Park, C. W. (1991). The differential role of characteristics of music on high- and low-involvement. *Journal of Consumer Research*, 18, 161-173.
- Moser, V. (2005). *Qualitätsmessung in öffentlichen Dienstleistungsorganisationen. Die Entwicklung eines Instruments zur Evaluierung von Musiktheateraufführungen* [Measuring quality in public service organizations. The development of an instrument to evaluate performances in music theater]. Unpublished master's thesis, University of Constance, Constance, Germany.
- Neuhoff, H. (2001). Die Altersstruktur von Konzertpublika. Querschnitte und Längsschnitte von Klassik bis Pop in kultursoziologischer Analyse [Age distribution of concert audience. Cross-sectional and longitudinal studies from classic to pop under a sociological focus]. *Musikforum*, 95, 64-83.
- Neuhoff, H., Boerner, S., & Renz, S. (2006). *Evaluation in music theatre*. Paper presented at the International Association of Empirical Aesthetics. Avignon.
- North, A. C., & Hargreaves, D. J. (1997). Music and consumer behavior. In D. J. Hargreaves (Ed.), *The social psychology of music* (pp. 268-289). New York and Oxford: Berghahn Books.
- Pavis, P. (1988). *Semiotik der Theaterrezeption* [Semiotics in the reception of theater]. Tübingen: Narr.
- Perkey, S. D. (1976). *Effects of positive and negative audience response on actors' nonverbal performance behaviour and on their attitudes*. Unpublished doctoral dissertation, Bowling Green State University, Ohio.
- Renz, S. (2006). *Evaluation in Non-Profit Organisationen. Eine empirische Untersuchung zur Rezeption der Aufführungsqualität im Musiktheater* [Evaluation in nonprofit organizations. An empirical investigation on the reception of performance quality in music theater]. Unpublished master's thesis University of Constance, Constance, Germany.
- Roth, P. L. (1994). Missing data: A conceptual review for applied psychologists. *Personnel Psychology*, 47, 537-560.
- Sagen, D. P. (1983). The development and validation of a university band performance rating scale. *Journal of Band Research*, 18, 1-11.
- Sauter, W., Kalvik A., & Isaksson, C. (1986). Das Publikum—Verständnis und Erlebnis von Theateraufführungen [The audience—Understanding and experience of theater performances]. In H. Schoenmakers (Ed.), *Performance theory. Advances in reception and audience research* (pp. 83-109). Amsterdam: Tijdschrift voor Theaterwetenschap.
- Schermelleh-Engel, K., & Moosbrugger, 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), 23-74.
- Tobias, S. (2004). Quality in the performing arts: Aggregating and rationalizing expert opinion. *Journal of Cultural Economics*, 28, 109-124.

- Wapnik, J., & Ekholm, E. (1997). Expert consensus in solo voice performance evaluation. *Journal of Voice*, 11(4), 429-436.
- Williamon, A., & Davidson, J. W. (2002). Exploring co-performer communication. *Musicae Scientiae*, 1(1), 53-72.

Direct reprint requests to:

Prof. Dr. Sabine Boerner
Universität Konstanz
Postfach D 88
Universitätsstr. 10, 78457
Konstanz, Germany
e-mail: Sabine.Boerner@uni-konstanz.de