

Small Influence of Performing from Memory on Audience Evaluation

REINHARD KOPIEZ[1]

Hanover University of Music, Drama and Media, Hanover, Germany

ANNA WOLF

Hanover University of Music, Drama and Media, Hanover, Germany

FRIEDRICH PLATZ

State University of Music and Performance, Stuttgart, Germany

ABSTRACT: This study investigates the influence of an actual music stand on the evaluation of a videotaped audio-visual solo instrumental performance. Previous research has provided evidence that the presence of a score or music stand (obstructing the audience's view of the performer) might negatively influence the evaluation of the performance. However, due to methodological ambiguities, results in previous studies cannot be regarded as definitive. Thus, we conducted a replication study of Williamon (1999) with better control over confounding variables (e.g., varying levels of technical proficiency in different conditions). A violoncello player performed two pieces for solo instrument: once with a music stand on stage (pretending to play from score) and once without. The level of technical proficiency was kept constant in both performance presentations by the use of a pre-recorded, well-rehearsed performance from memory. Audio tracks were synchronized with the performance movements in a playback paradigm. Based on the performance evaluations from a web-based experiment ($N = 471$ participants), we found a significant but small effect size for the main effect of performance presentation (with vs. without music stand) ($d = 0.23$). We conclude that the audience's appreciation of a particular performance from memory might be based on factors other than the objective performance quality.

Submitted 2016 June 12; accepted 2017 January 4.

KEYWORDS: *performance evaluation, memory, audio-visual, playing by heart*

PLAYING from memory has a long tradition in music performance (Chaffin, Demos, & Logan, 2016), and performance without a visible score or music stand is often regarded as a hallmark of virtuosity. In fact, such a performance is expected from musicians in competitions, examinations and recitals. However, as described in Bitzan (2010) and Ginsborg (2017, in press), playing from memory has been a controversial issue in the history of music performance. For example, in his treatise on the principles of violin playing from 1756, Leopold Mozart opposed performance from memory as he assumed that beginners would accustom themselves to playing at random without a score. Instead, Leopold Mozart suggests that unmelodious, distasteful pieces or fugal movements would compel pupils to read at sight (cf. Ginsborg, 2017, in press). Nevertheless, musicians did in fact begin performing from memory. Early public appreciation of this is documented in a critique of a recital given by the cellist Bernhard Romberg in 1822: "Romberg's great freedom in his element shows already in his appearance. Spurning the printed music as an aide-memoire, he takes his place, the magic instrument in his hands, and, without hiding himself behind a music stand, presents to the public the whole picture of a free, unrestricted ruler of the kingdom of tones." (cf. Ginsborg, 2017, in press, p. 7). Presumably, the pianist Clara Schumann was the first instrumentalist to systematically perform without a score in public at a recital in Berlin in 1837 (Mishra, 2010), where she played Beethoven's Piano Sonata, Op. 57 from memory at the age of 18. As such, Clara Schumann promoted the modern ideal of performance from memory (for a detailed discussion, see de Vries, 1996) rather than the pianist Hans von Bülow, who performed in London between 1873 and 1888 (see Scholes, 1947). To be more precise, according to the autobiographical memories of Clara's stepsister, Marie Wieck, it was Clara's father (and first piano teacher) Friedrich Wieck who introduced recitals from memory and encouraged her to train this performance style.



However, as reported by Marie Wieck (1912, p. 240), this was limited to solo repertoire. Piano concertos and chamber music were still performed from score. Only in the last years of the 19th century did performances from memory become a new standard. The public reaction to Clara's new approach was ambiguous: Although audiences must have been enthralled, the novelist Bettina von Arnim (Litzmann, 1925, p. 107) characterized Clara's behavior as "arrogant" (*mit welcher Prätension*) because she departed from the standard of performing from the score (*und nun ohne Noten!*). Until then, performance from memory was only allowed for free improvisation. At the same time, in his treatise on piano playing from 1839, the piano pedagogue Carl Czerny (1991, p. 70) developed a new ideal for the performer: Playing from memory was now regarded as an "honorable quality" (*ehrende Eigenschaft*) that offered "freedom and ease" (*Freiheit und Leichtigkeit*) and came as close as possible to the idea of free improvisation (*gleichsam schon einer freien Improvisation nahekommt*). In other words, performance from memory was meant to create the illusion that the soloist was a creative improviser and not a mere executor of a fixed composition. Consequently, Czerny recommended a repertoire of at least a dozen pieces ready to be performed from memory (for an overview of the history of playing from memory and memorization pedagogy, see Mishra, 2010, and Ginsborg, 2017, in press).

After much discussion among critics and musicians, Karl Schmidt addressed the pros and cons of playing or conducting from memory in his pioneering survey (1897a, 1897b, 1897c). Schmidt came to the conclusion that performance from memory is dispensable as it is merely time consuming, cannot be regarded as a special achievement of the performer, and only limits the size of a performer's repertoire. Responses to his essay from famous musicians, such as Richard Strauss, Carl Reinecke, and Joseph Rheinberger, contributed to the highly controversial discussion (for a more detailed discussion, see Wehmeyer, 1983, p. 173ff.). Similarly, the first survey on artistic attitudes toward playing from memory was initiated by the musicologist Wilhelm Altmann (1907a, 1907b). His question of whether artists should play from memory received responses from famous musicians, such as Ferruccio Busoni, across a broad range of opinions that provided no clear picture. Artists who supported performance without a score emphasized that such performances looked better, had more potential for interaction between musicians and the audience, and gave the interpreter more freedom for expressive gestures and movements. From a modern perspective, one could argue that the development towards the standard of playing from memory – at least for professional musicians – was likely the logical consequence of an increasing demand for "magic perfection" (Hughes, 1915). However, the performance practice of playing from memory depends on the musical genre (Ginsborg, 2004): One rarely sees folk, pop, or rock musicians entering the stage with a score, but as a general rule, chamber and symphonic music is played from notation (with only a few exceptions, such as the Kolisch Quartet, who performed avant-garde music by Schoenberg and Bartók without music stands, see *Kolisch Quartet*).

Surprisingly, only a small number of empirical studies address the challenging questions of how and if performance evaluation of classical music is influenced by the presence of a music stand. As Platz and Kopiez (2012) revealed in a meta-analysis, visual information can contribute significantly to the overall evaluation of an audio-visual performance by about 0.5 standard deviations. A visible score or music stand might be an important visual component of a live music performance, along with clothing (Griffiths, 2008, 2009), expressive movements and physical appearance (Davidson, 1993; Davidson & da Costa Coimbra, 2001; Tsay, 2013), attractiveness (Ryan & Costa-Giomi, 2004; Ryan, Wapnick, Lacaille, & Darrow, 2006), unconventional playing techniques (Lehmann & Kopiez, 2013), visible signs of engagement (Behne, 1994b), facial expressions (Thompson, Graham, & Russo, 2005; Thompson & Russo, 2007) and conductor gestures (Wöllner, 2008; Wöllner & Auhagen, 2008). In a study on the influence of first impressions on the audience's motivation to continue the performance evaluation, Platz and Kopiez (2013) showed that the visual cue of a score or music stand could very well have an influence on this process.

In a pioneering study, Williamon (1999) investigated the influence of playing from memory on the audience's audio-visual performance evaluation. The author tested whether performance "by heart" makes communication of expressive cues to the audience easier, as the performer can then move freely and is unconstrained by the visual contact with the music stand. Thus, audio-visual performances without a visible music stand should theoretically be evaluated more positively. This view is in line with Hughes' (1915, p. 595) argument in favor of playing from memory, arguing that this gives the performer the "absolute freedom of expression and the most direct psychological connection with the audience." To test presumed differences in the evaluation of memorized and non-memorized performances, Williamon (1999) produced in his study video recordings of three Preludes from Bach's *Suites for Violoncello Solo* (BWV 1007, 1008, and 1009) performed by a female cellist. In Condition 1 (the so-called "initial performance"), the cellist played each prelude as soon as she felt able to produce a satisfactory public performance. For this condition, she was asked not to memorize the pieces, and a music stand was used for the recording. In Conditions 2 and 3, recordings were produced as soon as the performer could play

from memory, which was the case about one month after the performance of Condition 1. In Condition 2, an empty music stand was visible, but the performer played from memory, and in Condition 3 no stand was visible. Condition 4 was matched to Condition 1 and again a music stand was used; however, it partially obstructed the audience's view. In Condition 5, the music stand was visible but the camera distance was increased to 5 meters. Random selections of video recordings (each participant rated at least one memorized performance) were evaluated by musicians and non-musicians in small groups on color television (28 inch, maximum distance = 8 meters). The following items were used for evaluation (6-point scale): overall quality, musical understanding (musicality), technical proficiency, and communicative ability. Ratings were analyzed for pieces and items separately.

As a result, combined ratings (all four performance evaluation scales) for the memorized conditions (Conditions 2 and 3) were significantly higher when compared to the non-memorized conditions (Conditions 1, 4 and 5). For Prelude 2 (BWV 1008), an interaction effect was observed: Musicians gave higher scores (combined ratings) for the memorized performances than non-musicians. The author concludes that "[ratings of] overall quality, musicality, technical proficiency and communication, when considered together, are improved by performing from memory" (Williamon, 1999, p. 89).

However, differences in evaluations between memorized and non-memorized performances can also be explained by certain confounding variables: As the author states, the performer might have benefitted from the extra month of practicing between the recordings of Condition 1 (non-memorized, visible music stand) and the later Condition 2 (memorized, visible music stand). Unfortunately, the original video recordings were lost (A. Williamon, personal communication, August 2014). The author concludes that, "These results suggest that the additional month of practice was, indeed, beneficial" (Williamon, 1999, p. 92). To summarize, neither the experimental design nor the data analysis of single performances and single evaluation items leads to generalizable conclusions.

Although we acknowledge the merits of this pioneering, but now outdated, study on the influence of visual cues on performance evaluation, there are many reasons that motivated us to start a replication:

- Based on the (analogue) video editing techniques of the 1990s, it was very difficult to avoid a confounding effect between the variables *Presentation Mode* and *Level of Technical Proficiency*. This would only have been possible by using dubbed versions with pre-recorded and synchronized audio tracks. Today, digital video editing and post production make it easy to hold audio tracks constant by means of carefully synchronized pre-recorded audio tracks.
- Ratings of the original study were analyzed with problematic statistical tools (invalidated sum score, multiple *t*-tests without reported alpha-correction, etc.). Today, increased computer power enables us to apply more advanced methods of statistical analysis, such as probabilistic test theory.
- The sample size of groups in the original study was small. For example, the group sizes in the repeated measures design varied between 4 and 16 participants (total: $N = 50$ musicians and $N = 36$ non-musicians). Today, experimental planning is determined by ideas on a priori test power (Ellis, 2010), which might favor fewer experimental conditions as well as a web-based experiment instead of a traditional lab-experiment.
- In the original study, no information on effect sizes was reported. Today, journals not only ask for statistical significance but also for relevance (standardized magnitude) of observed differences, which are best represented by effect sizes stated in the results section (APA, 2010).

Rationale of the Study

In our replication study, we reviewed Williamon's (1999) hypothesis that when players perform from memory, they receive a better evaluation from the audience than when they require the aid of a music stand. In addition, we assumed an interaction effect: In terms of the formation of an impression (Platz & Kopiez, 2013), we assumed that less musically sophisticated observers would be more impressed by the performance from memory and thus would give it a more positive evaluation.

METHOD

Design

A 2×2 between subjects design with the factors *Performance Presentation* (playing from memory vs. playing with a music stand) and *Participant's Degree of Musical Sophistication* (high vs. low) was used. The study was conceived as an online experiment (Reips, 2002; 2012). An a priori power analysis by means of the software G*Power V 3.1 (Faul, Erdfelder, Lang, & Buchner, 2007), with an assumed medium effect size of $f = 0.25$, $\alpha = .05$ and a test power of $1 - \beta = .90$ for the performance presentation \times degree of sophistication interaction, resulted in a total sample size of at least $N = 180$ participants, and a sample size of $N = 102$ participants for the main effect "condition of presentation" (t -test).

Stimuli

The beginnings of two works for violoncello solo by J. S. Bach (Prelude from BWV 1007 and Gigue from BWV 1009) were selected for the experiment, as they are part of a professional cellist's standard repertoire. Both pieces were well-rehearsed and memorized by an advanced male music student, who was a different cellist than in the original study by Williamon (1999). An audio track was pre-recorded in the same concert hall as used for the later video recording. The cellist obtained audio tracks of both pieces one week before the video recordings in order to practice a playback performance to his own recordings.

The video recording was conducted in a concert hall with professional background light. An HD video camera (Canon Legria HF G10) was used as the video recording device. The playback track was performed by in-ear monitoring with hidden earphones (masked by camera perspective) and a 2 bar count-in (the recording setup is shown in Figure 1a and b). Video recordings were produced with the same cellist in the same seat position for the conditions "with music stand" and "without music stand" (see Figure 2a and b). In the former condition, the music stand did not interfere with the view of the cellist or his instrument, unlike the music stand in Williamon's study (1999) that obstructed the view of the cellist in several conditions. The performer was instructed to look at the score from time to time during the condition "with music stand", although he played all pieces from memory. Multiple videos were recorded (with constant camera position medium shot) until the performer and experimenters were satisfied with the synchronization of the video and playback tracks. The final versions of audio-visual recordings were produced by the software Adobe Premiere (V 5.5), and fade-in/fade-out was added. The stimulus length was limited to about 45 seconds.

In a pre-test, a selection of 10 audio-visual recordings (two versions for each of the four conditions plus two control videos with an audio-visual asynchrony of 200 ms) were evaluated for synchronicity and persuasiveness (4-point Likert scale) by a group of 13 experts (music students) on a large screen (1.7×2 m). The four best synchronized video versions were selected for the final experiment (see Video S1 to S4 available at: <http://hdl.handle.net/1811/81127>).



Fig. 1a (left) and b (right). The left picture shows the setup for the video recording of both pieces. For better synchronization in the video post processing, the pre-recorded audio track with preceding click counter was sourced for the camera microphone via CD player. The right picture shows the preparation of the performer with an in-ear monitor fixed to his right ear. The monitor cable was adhered to his neck and was guided down his back such that all the technical equipment was invisible to the video observer.

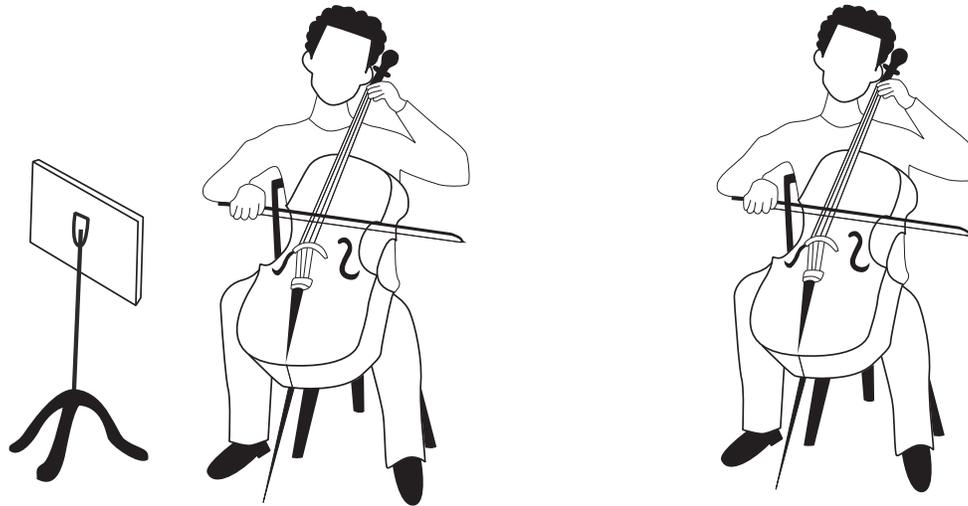


Fig. 2a (left) and b (right). View of the cellist during performance (a) with and (b) without music stand.

Participants

The study was conducted as an online experiment on the research platform *SoSci Survey* (<https://www.sosicisurvey.de>). In total, a convenience sample of $N = 471$ participants took part in the online study. Participants were acquired from mailing lists of the Hanover University of Music, Drama and Media (Germany) and from various Facebook groups. Degree of musical sophistication was assessed by self-reports on total years of private lessons and years of daily practice. Data collection occurred in the summer of 2014.

Procedure

Participants gave informed online consent. After adjustment of the loudness level, a technical inspection of the internet connection and the audio-visual setup, the experimenter showed the participants a practice video order to familiarize them with the rating procedure before they rated the test videos. To optimize the familiarization in the practice trial, this video was taken from the test stimuli and was different from the test video we used subsequently. Participants then watched the test video, again followed by the rating. The practice and test stimuli were always different pieces to avoid a direct comparison within the same piece. All other combinations (with/with, without/without, with/without, without/with music stand) were chosen randomly by the system. The entire duration for the study was about 15 minutes. All general recommendations for internet-based experimenting were considered (see Reips, 2002; 2012).

Items for Evaluation

We decided not to adopt the evaluation items from Williamon's original study for several reasons: (a) No theoretical justification for the selection of the four items (overall quality, musicality, technical proficiency and communication) was given in the original study; and (b) items such as musicality, overall quality and communication are most probably ambiguous and presumably multidimensional (at least for the communicative qualities of a performer; evidence for multidimensionality could be shown by Platz & Kopiez, 2013). To the best of our knowledge, there is currently no performance evaluation scale which meets the requirements of modern probabilistic test theory (e.g., IRT). Thus, in a first step, items from previous performance evaluation studies by Berlo, Lemert and Mertz (1969), McClaren (1985), Behne (1994a), Thompson and Williamon (2003), Wrigley and Emmerson (2013) and Lehmann and Kopiez (2013) were included, resulting in a total of 13 items (see Table 1). A 4-point Likert scale was used for the rating procedure ("I found the cellist's performance ..."; 1 = not at all, 4 = very much).

Table 1. Collection of Items for the Performance Evaluation Scale. Original German terms used in the survey are shown in brackets.

#	Item	Source
1	Concentrated [<i>konzentriert</i>]	Behne (1994b), Lehmann and Kopiez (2013)
2	Committed [<i>engagiert</i>]	Behne (1994b), Thompson and Williamon(2003)
3	Relaxed [<i>entspannt</i>]	McClaren (1985)
4	Stressed [<i>gestresst</i>]	Lehmann and Kopiez (2013)
5	Authentic [<i>authentisch</i>]	Lehmann and Kopiez (2013)
6	Certain/confident [<i>sicher</i>]	Berlo (1969), Behne (1994a), Behne and Wöllner (2011), Ambady and Rosenthal (1993), Wrigley and Emmerson (2013)
7	Expressive [<i>ausdrucksvoll</i>]	Behne (1994a), Thompson and Williamon (2003), Behne and Wöllner (2011)
8	Empathetic [<i>empfindend</i>]	Behne (1994b), Ambady and Rosenthal (1993)
9	Rousing/enthusiastic [<i>mitreißend</i>]	Behne (1994b), Ambady and Rosenthal (1993)
10	Precise [<i>präzise</i>]	Behne (1994a, 1994b), Behne and Wöllner (2011)
11	Sonorous/resonant [<i>klangvoll</i>]	Behne (1994a, 1994b), Behne and Wöllner (2011)
12	Persuasive [<i>überzeugend</i>]	Behne (1994a), Behne and Wöllner (2011)
13	Professional [<i>professionell</i>]	Ambady and Rosenthal (1993)

RESULTS

Data Analysis

DATA PREPARATION

The complete process of data preparation and analysis is shown as a flow chart in Figure 3. In the first step, the total data set ($N = 471$) was filtered by removing incomplete data sets or those with implausible processing times ($n = 94$). The resulting sample was comprised of $n = 139$ males and $n = 238$ females.

INCREASING CONTRASTS BETWEEN GROUPS OF PARTICIPANTS

To increase the contrast between groups of high vs. low degrees of musical sophistication, in the second step, we excluded $n = 47$ participants who were neither musical experts (≥ 8 years of music lessons, $n = 238$) nor amateurs (0-4 years of music lessons, $n = 92$). A final sample of $n = 330$ cases remained for statistical analysis, which can be seen in Table 2. This procedure of contrast enhancement was also used in previous research (e.g., Witek, Clarke, Wallentin, Kringelbach, & Vuust, 2014).

Table 2. Description of the Sample ($n = 330$ valid cases).

Variable	<i>M</i>	<i>SD</i>
Whole sample: Age	28.6	10.1
Experts: Years of private instrumental lessons	13.6	4.1
Experts: Years of daily practice	10.6	7.5
Amateurs: Years of private instrumental lessons	1.7	1.6
Amateurs: Years of daily practice	1.8	3.7

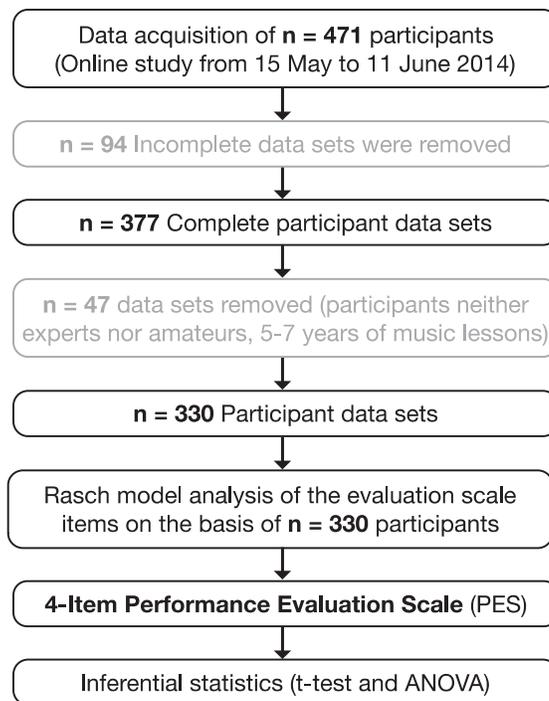


Fig. 3. Flow chart of the data preparation process.

ITEM ANALYSIS

In the third step, using the remaining $n = 330$ data sets, we dichotomized and analyzed the 13 scale items with respect to their unidimensionality as defined by the IPL model (or Rasch model, see De Ayala, 2009). The idea behind this model is that the participant's answer to each item is probabilistically determined only by his or her mindset and by chance, leading to a *unidimensional measuring model*. Latent parameters for the participants and items are estimated and represented on a *ratio scale*. In addition, the items are *measurement invariant* making it irrelevant which set of (the model-fitting) items is tested to measure a participant's mindset towards the stimuli (see also Gregory, 2015). However, before these (and several other) characteristics can be assumed, the items need to be statistically validated as to whether they comply with the Rasch model. Such an analysis was performed as described in Koller and Hatzinger (2013) and Fischer and Molenaar (1995) and implemented by Mair, Hatzinger, Maier, and Rusch (2016). A set of parametric and non-parametric model tests were applied, which focused on unidimensionality, local independence of the items, parallel and strictly increasing item characteristic curves (ICC), and specific objectivity.

After this iterative process of item selection and dismissal, four items fulfilled the criteria of the Rasch model (*committed*, *authentic*, *certain/confident* and *rousing/enthusiastic*) and thus constitute the *Performance Evaluation Scale* (PES; see Table S1 in the Supplementary Material Online document available at: <http://hdl.handle.net/1811/81127> for detailed information on the statistical procedures).

PERFORMANCE EVALUATION SCORE

In the fourth and final step of data preparation, the dichotomized marks for the four remaining items from IRT analysis (*committed*, *authentic*, *certain/confident* and *rousing/enthusiastic*) were summed up so that the resulting Performance Evaluation Scale (PES) was within a score range from 0 (no agreement with any item) to 4 (agreement with all items).

Differences between Main Conditions

Between groups comparisons of PES score revealed a small but significant advantage in favor of the presentation without music stand (with music stand: $M = 2.95$, $SE = 0.10$, $n = 167$; without music stand: $M = 3.21$, $SE = 0.08$, $n = 163$; $t(310.55) = -2.08$, $p = .02$ [one-tailed], $d = 0.23$). As shown in Figure 4, the difference between groups was less than 0.3 scale steps.

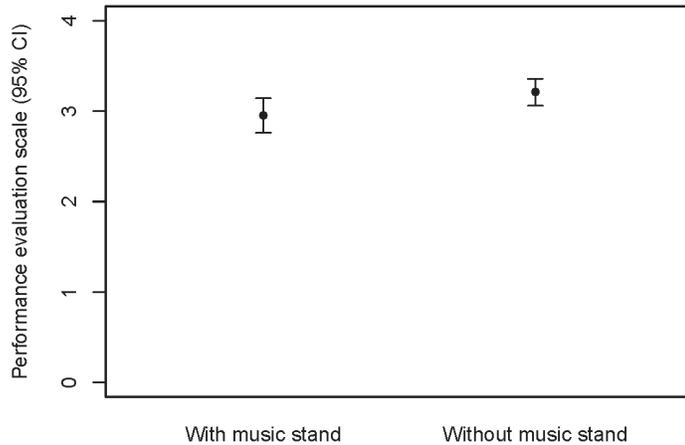


Fig. 4. Mean values of the Performance Evaluation Scale for the conditions with vs. without music stand (PES score: 0 = poor, 4 = very good). In both conditions the performer played the piece from memory, but in the condition with the visible music stand, he looked at the score. The data were averaged across both pieces.

Interaction Effects

INFLUENCE OF MUSICAL BACKGROUND

Although we had hypothesized that less sophisticated participants (musical amateurs) would provide higher evaluations in the condition without music stands, the error bar diagram (Figure 5) shows a different result: The interaction between Performance Presentation \times Degree of Sophistication revealed only small differences between groups, which were not statistically significant (see Table 3 for descriptive statistics).

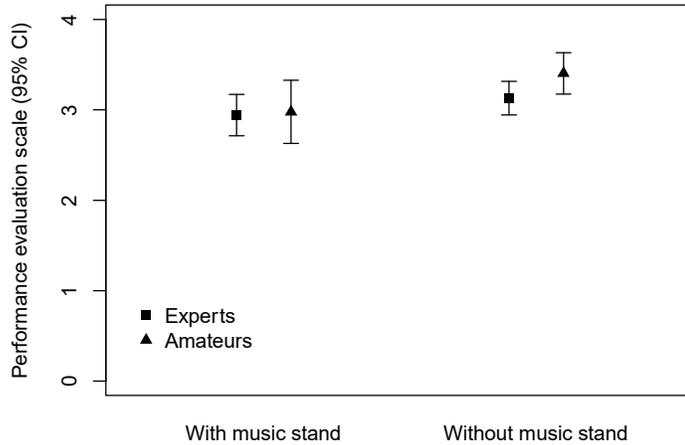


Fig. 5. Mean values of the Performance Evaluation Scale for performance presentation with vs. without music stand and high (experts with ≥ 8 years of music lessons) vs. low (amateurs with 0-4 years of music lessons) degree of musical sophistication (PES score: 0 = poor, 4 = very good).

Table 3. Descriptive Statistics for the Performance Presentation \times Degree of Sophistication Interaction.

	With music stand		Without music stand	
	Experts	Amateurs	Experts	Amateurs
<i>M</i>	2.94	2.98	3.13	3.40
<i>SE</i>	0.12	0.18	0.09	0.12
<i>n</i>	122	45	116	47

Note. Results from the ANOVA for the interaction effect: $F(1,326) = 0.75, p = .39, \eta^2 = 0.002$.

DISCUSSION AND CONCLUSIONS

We conclude that the audience's appreciation of a particular performance from memory (without visible music stand) might be based on factors other than the objective performance quality. Additionally, although performance from a score is often confounded with a player's supposed lack of technical proficiency – which was the case in Williamon's (1999) study – the mere use of a music stand may have only a marginal influence on performance evaluation. As previous research has shown, evaluation processes result from a complex interplay of variables. Thus, no single variable has a dominant influence. In a meta-analysis, Platz and Kopiez (2012) showed that all visual components of stage behavior influenced the evaluation of a performance with an effect size of $d = 0.51$. Moreover, in a study on the features of persuasiveness, Platz and Kopiez (2013) revealed that the formation of the audience's impression is determined by a bundle of factors, such as gaze direction, stance width or step size of the performer. According to the authors, the key construct that best describes this complex evaluative process is that of "appropriateness": It is only when there is a match between the audience's expectations and the performer's stage behavior that the audience wishes for an ongoing performance. As long as it remains unclear as to why we should even feel that a music stand interferes with the performance, this working tool might have only a small influence on our evaluation of a performance.

Furthermore, we could not find a significant difference in the evaluations of musical experts and amateurs and had to dismiss our hypothesis that less musically sophisticated observers would be more impressed by the performer's playing from memory. As a by-product of our study, the development of a 4-item Performance Evaluation Scale might be a sustainable contribution to the future elaboration of a more comprehensive evaluation tool. To the best of our knowledge, there is currently no inventory for the evaluation of music performance that is compatible with the highest standards of probabilistic test theory. The suggested items of the PES might be a first step in this direction.

END MATTER

Acknowledgements

We thank Cosimo Carovani for his patience and cooperation during the recording of the videos, Theresa Tamoszus for her support during the implementation of the study, and Maria Lehmann for the careful language editing.

Supplemental Material

Supplemental Online Material (video recordings etc.) are available at: <http://hdl.handle.net/1811/81127>

NOTES

[1] Correspondence can be addressed to: Reinhard Kopiez, Hanover University of Music, Drama and Media, Emmichplatz 1, 30175 Hanover, Germany, reinhard.kopiez@hmtm-hannover.de.

REFERENCES

- Altmann, W. (1907a). Sollen die Künstler auswendig spielen? Eine Anregung [Should artists play from memory? A suggestion]. *Die Musik*, 6(22), 284-285.
- Altmann, W. (1907b). Sollen die Künstler auswendig spielen? Zwei Erwiderungen [Should artists play from memory? Two responses]. *Die Musik*, 6(23), 146-151.
- Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology*, 64(3), 431-441. <https://doi.org/10.1037/0022-3514.64.3.431>
- APA (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: American Psychological Association.
- Behne, K.-E. (1994a). Blicken Sie auf die Pianisten?! - Zur bildbeeinflussten Beurteilung von Klaviermusik im Fernsehen ['Do you watch the pianists?!' Visual influences on judgments of piano music on television]. In K.-E. Behne (Ed.), *Gehört - Gedacht - Gesehen. Zehn Aufsätze zum visuellen, kreativen und theoretischen Umgang mit Musik* [Heard, thought, seen: Ten essays on visual, creative, and theoretical dealings with music] (pp. 9-22). Regensburg: Bosse.
- Behne, K.-E. (1994b). Schönheit oder Engagement? Über die notwendigen visuellen Attribute eines Musikers [Beauty or engagement? On a musician's necessary visual attributes]. In K.-E. Behne (Ed.), *Gehört - Gedacht - Gesehen* [Heard, thought, seen: Ten essays on visual, creative, and theoretical dealings with music] (pp. 47-70). Regensburg, Germany: Bosse.
- Behne, K.-E., & Wöllner, C. (2011). Seeing or hearing the pianists? A synopsis of an early audiovisual perception experiment and a replication. *Musicae Scientiae*, 15(3), 324-342. <https://doi.org/10.1177/1029864911410955>
- Berlo, D. K., Lemert, J. B., & Mertz, R. J. (1969). Dimensions for evaluating the acceptability of message sources. *Public Opinion Quarterly*, 33(4), 563-576. <https://doi.org/10.1086/267745>
- Bitzan, W. (2010) *Auswendig lernen und spielen: Über das Memorieren in der Musik* [Learning and playing by heart: Memorizing music]. Frankfurt a.M., Germany: Peter Lang.
- Chaffin, R., Demos, A. P., & Logan, T. (2016). Performing from memory. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (2nd ed., pp. 559-571). Oxford: Oxford University Press.

Czerny, C. (1991). *Von dem Vortrage: Dritter Teil der Vollständigen theoretisch-practischen Pianoforte-Schule op. 500* [On performance: Third part of the theoretical-practical piano-forte textbook Op. 500] (Original work published 1839). Wiesbaden, Germany: Breitkopf & Härtel.

Davidson, J. W. (1993). Visual perception of performance manner in the movements of solo musicians. *Psychology of Music, 21*(2), 103-113. <https://doi.org/10.1177/030573569302100201>

Davidson, J. W., & da Costa Coimbra, D. (2001). Investigating performance evaluation by assessors of singers in a music college setting. *Musicae Scientiae, 5*(1), 33-54. <https://doi.org/10.1177/102986490100500103>

De Ayala, R. J. (2009). *The theory and practice of item response theory*. New York: Guilford Press.

de Vries, C. (1996). *Die Pianistin Clara Wieck-Schumann: Interpretation im Spannungsfeld von Tradition und Individualität* [The pianist Clara Wieck-Schumann: Interpretation within the conflict between tradition and individuality]. Mainz, Germany: Schott.

Ellis, P. D. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511761676>

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191. <https://doi.org/10.3758/BF03193146>

Fischer, G. H., & Molenaar, I. W. (1995). *Rasch models: Foundations, recent developments, and applications*. New York: Springer. <https://doi.org/10.1007/978-1-4612-4230-7>

Ginsborg, J. (2004). Strategies for memorizing music. In A. Williamon (Ed.), *Musical excellence: Strategies and techniques to enhance musical performance* (pp. 123-141). Oxford: Oxford University Press.

Ginsborg, J. (2017, in press). Memory in music listening and performance. In P. Hansen & B. Blaesing (Eds.), *Performing the remembered present: The cognition of memory in dance, theatre and music*. London: Bloomsbury.

Gregory, R. J. (2015). *Psychological testing: History, principles, and applications*. Boston, MA: Pearson.

Griffiths, N. K. (2008). The effects of concert dress and physical appearance on perceptions of female solo performers. *Musicae Scientiae, 12*(2), 273-290. <https://doi.org/10.1177/102986490801200205>

Griffiths, N. K. (2009). 'Posh music should equal posh dress': An investigation into the concert dress and physical appearance of female soloists. *Psychology of Music, 38*(2), 159-177. <https://doi.org/10.1177/0305735608100372>

Hughes, E. (1915). Musical memory in piano playing and piano study. *The Musical Quarterly, 1*(4), 592-603. <https://doi.org/10.1093/mq/I.4.592>

Kolisch Quartet. (n.d.). In *Wikipedia*. Retrieved 10 May 2016, from https://en.wikipedia.org/wiki/Kolisch_Quartet.

Koller, I., & Hatzinger, R. (2013). Nonparametric tests for the Rasch model: Explanation, development, and application of quasi-exact tests for small samples. *InterStat*(21, November), Retrieved 20 March 2016, from <http://interstat.statjournals.net/YEAR/2013/articles/1311002.pdf>.

Lehmann, M., & Kopiez, R. (2013). The influence of on-stage behavior on the subjective evaluation of rock guitar performances. *Musicae Scientiae, 17*(4), 472-494. <https://doi.org/10.1177/1029864913493922>

- Litzmann, B. (1925). *Clara Schumann. Ein Künstlerleben nach Tagebüchern und Briefen* [Clara Schumann: An artists's life in diaries and letters] (6th ed. Vol. 1). Leipzig: Breitkopf & Härtel.
- Mair, P., Hatzinger, R., Maier, M., & Rusch, T. (2016). Package 'eRm' [Computer software manual]. Retrieved 21 March 2016, from <http://cran.r-project.org/web/packages/eRm/eRm.pdf>
- McClaren, C. A. (1985). *The influence of visual attributes of solo marimbists on perceived qualitative response of listeners*. Doctoral dissertation, The University of Oklahoma, USA. Retrieved from <http://search.proquest.com/docview/303414658?accountid=16198>
- Mishra, J. (2010). A century of memorization pedagogy. *Journal of Historical Research in Music Education*, 32(1), 3-18. <https://doi.org/10.1177/153660061003200102>
- Platz, F., & Kopiez, R. (2012). When the eye listens: A meta-analysis of how audio-visual presentation enhances the appreciation of music performance. *Music Perception*, 30(1), 71-83. <https://doi.org/10.1525/mp.2012.30.1.71>
- Platz, F., & Kopiez, R. (2013). When the first impression counts: Music performers, audience, and the evaluation of stage entrance behavior. *Musicae Scientiae*, 17(2), 167-197. <https://doi.org/10.1177/1029864913486369>
- Reips, U.-D. (2002). Standards for internet-based experimenting. *Experimental Psychology*, 49(4), 243-256. <https://doi.org/10.1026//1618-3169.49.4.243>
- Reips, U.-D. (2012). Using the internet to collect data. In H. Cooper (Ed.), *APA handbook of research methods in psychology* (Vol. 2 - Research designs, pp. 291-310). Washington, DC: American Psychological Association. <https://doi.org/10.1037/13620-017>
- Ryan, C., & Costa-Giomi, E. (2004). Attractiveness bias in the evaluation of young pianists' performances. *Journal of Research in Music Education*, 52(2), 141-154. <https://doi.org/10.2307/3345436>
- Ryan, C., Wapnick, J., Lacaille, N., & Darrow, A.-A. (2006). The effects of various physical characteristics of high-level performers on adjudicators' performance ratings. *Psychology of Music*, 34(4), 559-572. <https://doi.org/10.1177/0305735606068106>
- Schmidt, K. (1897a). Auswendigspielen und Auswendigdirigieren - Fortsetzung [Playing from memory and conducting from memory - A continuation]. *Centralblatt für Instrumentalmusik, Solo- und Chorgesang*, 12(2), 28-31.
- Schmidt, K. (1897b). Auswendigspielen und Auswendigdirigieren - Schluss [Playing from memory and conducting from memory - Conclusion]. *Centralblatt für Instrumentalmusik, Solo- und Chorgesang*, 12(3), 53-54.
- Schmidt, K. (1897c). Auswendigspielen und Auswendigdirigieren [Playing from memory and conducting from memory]. *Centralblatt für Instrumentalmusik, Solo- und Chorgesang*, 12(1), 2-5.
- Scholes, P. A. (1947). *A century of musical life in Britain as reflected in the pages of the Musical Times*. Freeport, NY: Books for Libraries Press.
- Thompson, S., & Williamon, A. (2003). Evaluating evaluation: Musical performance assessment as a research tool. *Music Perception*, 21(1), 21-41. <https://doi.org/10.1525/mp.2003.21.1.21>
- Thompson, W. F., Graham, P., & Russo, F. A. (2005). Seeing music performance: Visual influences on perception and experience. *Semiotica*, 156(1/4), 203-227. <https://doi.org/10.1515/semi.2005.2005.156.203>
- Thompson, W. F., & Russo, F. A. (2007). Facing the music. *Psychological Science*, 18(9), 756-757. <https://doi.org/10.1111/j.1467-9280.2007.01973.x>

- Tsay, C.-J. (2013). Sight over sound in the judgment of music performance. *PNAS*, *110*(36), 14580-14585. <https://doi.org/10.1073/pnas.1221454110>
- Wehmeyer, G. (1983). *Carl Czerny und die Einzelhaft am Klavier* [Carl Czerny and solitary confinement at the keyboard]. Kassel, Germany: Bärenreiter.
- Wieck, M. (1912). *Aus dem Kreise Wieck-Schumann* [From Wieck-Schumann's circle of friends]. Dresden, Germany: Pierson.
- Williamon, A. (1999). The value of performing from memory. *Psychology of Music*, *27*(1), 84-95. <https://doi.org/10.1177/0305735699271008>
- Witek, M. A. G., Clarke, E. F., Wallentin, M., Kringelbach, M. L., & Vuust, P. (2014). Syncopation, body-movement and pleasure in groove music. *PLoS One*, *9*(4), e94446. <https://doi.org/10.1371/journal.pone.0094446>
- Wöllner, C. (2008). Which part of the conductor's body conveys most expressive information? A spatial occlusion approach. *Musicae Scientiae*, *12*(2), 249-272. <https://doi.org/10.1177/102986490801200204>
- Wöllner, C., & Auhagen, W. (2008). Perceiving conductors' expressive gestures from different visual perspectives. An exploratory continuous response study. *Music Perception*, *26*(2), 129-143. <https://doi.org/10.1525/mp.2008.26.2.129>
- Wrigley, W. J., & Emmerson, S. B. (2013). Ecological development and validation of a music performance rating scale for five instrument families. *Psychology of Music*, *41*(1), 97-118. <https://doi.org/10.1177/0305735611418552>