

Assessing children's voices using Hornbach and Taggart's (2005) rubric

Andreas C. Lehmann¹ & Johannes Hasselhorn²

Hochschule für Musik Würzburg, Germany

¹ac.lehmann@hfm-wuerzburg.de, ²johannes.hasselhorn@hfm-wuerzburg.de

ABSTRACT

Background

Assessment of voice quality and performance is notoriously difficult, and even professional singers may not always agree on the quality of a voice or performance (Mitchell & Kenny, 2008). Although there is a mild consensus about what constitutes a good professional voice, untrained voices pose a serious challenge to raters. Educators have developed different types of assessment instruments (Salvator, 2010). For instance, Hornbach & Taggart (2005) used a researcher-developed, 5-point assessment rubric to assess elementary-age children (H&T rubric). Although satisfying interjudge reliabilities are reported by the authors of the rubric ($r = .76$ to $.97$), it is unclear what specific aspects of performance influence this overall (summative) impression.

Aims

The goal of this study was to find a way of assessing children's vocal performances more objectively. More specifically, the aim was to study the reliabilities of experts' ratings of children's performances (using the summative H&T rubric), and to predict such summative, overall judgments of singing achievement from more detailed performance ratings.

Method

55 children with an average age of 11.1 ($SD = .79$; Mode = 11) were individually asked to sing a supposedly well-known song (the German national anthem) and another song they had practiced previously in school. Performances were recorded.

Subsequently, three independent expert judges rated the performances. The raters were one singer who also teaches voice and one public school music teacher. Both had studied voice at a conservatory. A third rater, a public school teacher without specialization in music, had to be excluded in the course of the analysis due to low agreement with the other two raters.

Ratings consisted of (a) 19 five-point rating scales regarding specific aspects (e.g., articulation, matching of given starting notes, rhythm) and (b) a comprehensively, summative evaluation using the H&T rubric.

Results

Averaged expert ratings on the H&T rubric for the two songs were significantly correlated, $r = .80$, $p < .01$. The mean ratings for the two songs were not significantly different, suggesting that both songs were sung equally well, and vocal performance was similar on both occasions. Subsequent analyses were only conducted on data from German national anthem renditions.

Inter-rater consistency was estimated using intra-class correlations ($ICC[2,1]$) to establish which aspects of performance were most reliably rated by expert raters. In

general, intra-class correlations are used to quantify agreement of raters with regard to groups of observations. Here, we used a two-way, random effect ICC, because two randomly selected assessors both rated all vocal performances. ICCs varied between .06 and .81 (see Table 1).

Table 1
Intra-class Correlations ($ICC[2,1]$) for Two Experts' Ratings of Children's Vocal Performance

Rated vocal performance attributes	ICC
Piano starting tone: match – no match	.81
Type of performance: speechlike – singing	.77
Melody execution: secure – insecure	.67
Attitude: secure – insecure	.65
Loudness: soft – loud	.65
Voice-ear coordination: fitting – not fitting	.65
Tessitura: small – age appropriate	.64
Text integration: fluent – stumbling	.63
Interpretation: differentiated – undifferentiated	.56
Hornbach & Taggart Scoring Rubric (overall)	.80

Note. Only ICCs $> .55$ shown. All F-tests $F(54) > 3.7$, $p < .01$.

We created a scale using eight of the individual attributes listed in Table 1 (excluding "Loudness" for technical reasons; it may have been affected by the singer's distance from the microphone), which resulted in a scale with a reliability of .94 (Cronbach's alpha).

This newly created scale of specific performance features was then related to the overall measure (H&T rubric), resulting in a correlation of .93, $p < .01$. Thus, the aggregated individual features of children's sung performance were highly related to the summative expert ratings.

The final question addressed the relative importance of the individual scale items in predicting our overall measure (H&T rubric). In a stepwise linear regression analysis we entered all scale items as predictors and H&T rubric as the criterion. Two variables, namely "Melody execution" and "Piano starting tone" entered the equation, explaining a total of 92 percent (adjusted) of the variance (see Table 2).

Table 2
Regression Analysis Summary for Specific Vocal Performance Characteristics Predicting a Summative Rating (H&T rubric)

Variable	β	SE β	stand. β	ΔR^2
Melody execution	.88	.06	.77	.89
Piano starting tone	.23	.05	.25	.03

Note. $R^2 = .92$ (adjusted).

Conclusions

Summative measures, here the Hornbach and Taggart (2005) rubric (H&T rubric), can be seen as effective assessment instruments when used by experts. The ratings on this rubric could be strongly predicted by a scale of individual aspects of performance ($R^2 = .92$). Using the H&T rubric may therefore be a parsimonious way to assess children's sung performances in a school setting.

However, it is unclear whether non-professionals are able to provide such reliable ratings of children's vocal performance. In our study, we had to exclude a non-music teacher due to lack of reliability (see Method section). More research will be necessary to corroborate this finding.

Currently, a replication of our study is under way, and additional data will likely be available at the time of the conference.

Ultimately, this research contributes to the goal of developing a computer-assisted scoring of vocal performances (see ICMPC12 contribution by Dittmar et al., "Automatic Singing Assessment of Pupil Performances"). Such a scoring would provide a valuable measure of performance that goes beyond the simple note matching scores of karaoke-type games.

Keywords

Singing, Children, Measuring Performance, Reliability of Judges

REFERENCES

- Hornbach, C. M., & Taggart, C. C. (2005). The relationship between developmental tonal aptitude and singing achievement among kindergarten, first-, second-, and third-grade students. *Journal of Research in Music Education*, 53, 322-331.
- Mitchell, H. F., & Kenny, D. T. (2008). Open Throat: Acoustic and Perceptual Support for Pedagogic Practice. *Journal of Singing*, 64(1), 429-441.
- Salvator, K. (2010). How Can Elementary Teachers Measure Singing Voice Achievement? A Critical Review of Assessments, 1994-2009. *Update: Applications of Research in Music Education*, 29(1), 40-47.