A New Speech, Spatial, and Qualities of Hearing Scale Short-Form: Factor, Cluster, and Comparative Analyses

Annie Moulin,[1,2] Judith Vergne,[1,2] Stéphane Gallego,[3,4,5] and Christophe Micheyl[6,7,8]

Objectives: The objective of this work was to build a 15-item short-form of the Speech Spatial and Qualities of Hearing Scale (SSQ) that maintains the three-factor structure of the full form, using a data-driven approach consistent with internationally recognized procedures for short-form building. This included the validation of the new short-form on an independent sample and an in-depth, comparative analysis of all existing, full and short SSQ forms.

Design: Data from a previous study involving 98 normal-hearing (NH) individuals and 196 people with hearing impairments (HI), non hearing aid wearers, along with results from several other published SSQ studies, were used for developing the short-form. Data from a new and independent sample of 35 NH and 88 HI hearing aid wearers were used to validate the new short-form. Factor and hierarchical cluster analyses were used to check the factor structure and internal consistency of the new short-form. In addition, the new short-form was compared with all other SSQ forms, including the full SSQ, the German SSQ15, the SSQ12, and the SSQ5. Construct validity was further assessed by testing statistical relationships between scores and audiometric factors, including pure-tone threshold averages (PTAs) and left/right PTA asymmetry. Receiver-operating characteristic analyses were used to compare the ability of different SSQ forms to discriminate between NH and HI (HI non hearing aid wearers and HI hearing aid wearers) individuals.

Results: Compared all other SSQ forms, including the full SSQ, the new short-form showed negligible cross-loading across the three main subscales and greater discriminatory power between NH and HI subjects (as indicated by a larger area under the receiver-operating characteristic curve), as well as between the main subscales (especially Speech and Qualities). Moreover, the new, 5-item Spatial subscale showed increased sensitivity to left/right PTA asymmetry. Very good internal consistency and homogeneity and high correlations with the SSQ were obtained for all short-forms.

Conclusions: While maintaining the three-factor structure of the full SSQ, and exceeding the latter in terms of construct validity and sensitivity to audiometric variables, the new 15-item SSQ affords a substantial reduction in the number of items and, thus, in test time. Based on overall scores, Speech subscores, or Spatial subscores, but not Qualities subscores, the 15-item SSQ appears to be more sensitive to differences in self-evaluated hearing abilities between NH and HI subjects than the full SSQ.

[1]INSERM U1028, CNRS UMR 5292, Lyon Neuroscience Research Center, Brain Dynamics and Cognition Team, University of Lyon, Lyon, France; [2]Université C. Bernard Lyon 1, Lyon Neuroscience Research Center, Brain Dynamics and Cognition Team, University of Lyon, Lyon, France; [3]Institut des Sciences et Technologies de Réadaptation (ISTR), University of Lyon, Lyon, France; [4]Audition Conseil, Lyon, France; [5]Laboratory of Integrative and Adaptive Neurosciences (LNIA) UMR 7260, Aix-Marseille University-CNRS, Marseille, France; [6]Starkey Hearing Technologies, Créteil, France; [7]Starkey Hearing Research Center, Berkeley, California, USA; and [8]INSERM U1028, CNRS UMR 5292, Cognition and Auditory Perception Team (CAP), Lyon Neuroscience Research Center, University of Lyon, Lyon, France.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and text of this article on the journal's Web site (www.ear-hearing.com).

Key words: Hearing disability, Hearing loss, Self-report measure, Short-form building, Spatial hearing, Speech.

(Ear & Hearing 2019;40;938–950)

INTRODUCTION

Self-report outcome measures have become an essential component in the evaluation of rehabilitation benefits for patients. In the field of audiology, several scales have been developed for the self-evaluation of various aspects of hearing, such as speech perception, binaural hearing, or hearing in challenging situations. Among these, the Speech Spatial and Qualities of Hearing Scale (SSQ) designed by Gatehouse and Noble (2004) is a widely used self-report measure of hearing and has already been translated into several languages (e.g., Dutch (Demeester et al. 2012), Korean (Kim et al. 2017), German (Kiessling et al. 2011), French and Portuguese (Gonsalez et al. 2015).

In its full form, the SSQ includes 49 items. One advantage of this relatively large number of items is that it makes it possible to explore some very specific aspects of hearing via three main subscales (Speech, Spatial, and Qualities) and 10 pragmatic subscales (Gatehouse & Akeroyd 2006). One disadvantage relates to the substantial amount of time and effort required for completion. Using a French version of the SSQ, Moulin et al. (2015) found that the self-assessed time to complete the scale in a hearing-impaired (HI) group varied from 10 min to 1 hr, with more than 25% of respondents reporting completion times above 25 min. This makes it difficult to use the SSQ in routine clinical use, or for swift assessments of hearing. Moreover, the number of missing responses tended to increase over the last third of the scale, with more missing responses for the most difficult-to-read items (Moulin et al. 2015). Indeed, with over a thousand words, the SSQ can be quite taxing for respondents.

In order to address this problem, short-forms of the SSQ have been developed. Demeester et al. (2012) created a short-form with five items (SSQ5) out of the Dutch version of the SSQ, specifically for screening purposes. Using the UK version of the SSQ, Noble et al. (2013) created a short-form with 12 items (SSQ12) based on a large multi-center data set. Neither of these short-forms include the main three subscales. Indeed, the aim of Noble et al. was "to compile a set of items that represent the scale as a whole," and that reflect the 10 pragmatic subscales of the SSQ, as defined by Gatehouse and Akeroyd (2006). Validating a German version of the SSQ, Kiessling et al. (2011) proposed a short-form containing 15 items, with five items per main subscale. The resulting short-form has the potential to retain the information present in the three main subscales of the SSQ. However, each of these short-forms was constructed based on results from a single data sample, an approach which should be avoided if possible,

<zdoi; 10.1097/AUD.0000000000000675>

because it tends to overestimate the expected reliability of the short-form (Putnam & Rothbart 2006), and fails to take into account potential variations of the optimal subset of items across different data samples (Widaman et al. 2011). To our knowledge, except for the SSQ5, which was built specifically for screening purposes, none of these short-forms have been validated yet on a different sample of participants from the one used to build it, which is one of the last recommended steps in short-form building. Indeed, the difficulty of short-form building is the trade-off between the potential gain in time to administer the scale and the potential loss in psychometric properties due to items reduction. This trade-off renders an independent assessment of the reliability and validity of the new measure essential (Smith et al. 2000).

In this context, the main goals of the present study were as follows. First, the aim was to create a short-form of the SSQ using a data-driven approach, taking into account SSQ results from as many studies as available, provided that the results were given per item, and following the recommended guidelines for short-form building (Smith et al. 2000; Stanton et al. 2002; Widaman et al. 2011). Second, this work allowed us to evaluate the performances of the preexisting and newly created short-forms on a new, independent sample of participants. To this aim, we tried to follow the recommendations of Smith et al. (2000) for avoiding the "sins of short-form development." Hence, the short-form was built using a data-driven approach and strategies specific to short-form building (Stanton et al. 2002), using published data from five SSQ studies involving four different languages (Dutch, English, French, and German). First, responses to items of the newly built short-form were extracted from existing data from a previous study in which the full SSQ was administered to 196 HI participants (non hearing aid wearers [HIN]) and 100 normal-hearing (NH) participants (Moulin et al. 2015). Subsequently, both the full SSQ and the newly built short-form were administered to a new sample of 88 HI subjects (all hearing aid wearers [HIHA]) and 35 NH subjects. In-depth analyses of all existing, full and short SSQ forms were performed. In particular, factor and hierarchical cluster analyses were used to check the factor structure and internal consistency of the new short-form. In addition, the new short-form was compared with all other SSQ forms, including the full SSQ, the German SSQ15, the SSQ12, and the SSQ5. Construct validity refers to how well the new scales cover the same content as the original full scale and relates to external factors the same way as the original scale. With reliability, it is the most important factor to respect when creating a short-form (John & Soto 2009). This was assessed by analyzing statistical relationships between scores and patients' characteristics, such as age and audiometric factors, including pure-tone threshold averages (PTAs) and left/right PTA asymmetry. In addition, the ability of the SSQ and all the short-forms to discriminate between NH and HI (HIN and HIHA) individuals was compared using receiver-operating characteristic (ROC) analyses. Those analyses consist of building curves that show the sensitivity (ability to detect impairment) as a function of the specificity (proportion of correctly recognized NH), as the cutoff score of the SSQ forms is moved across its range. This gives a summary statistic (the area under the curve) allowing to compare the quality of the discrimination between impairment and non-impairment.

MATERIALS AND METHODS

Short-Form Creation

In order to create the new short-form, we considered the following sources of information:

- Missing response rates across items for the German SSQ (Kiessling et al. 2011), the English SSQ (Akeroyd et al. 2014), and the French SSQ (Moulin et al. 2015; Moulin & Richard 2016a, 2016b).
- Results and normative values obtained per item (mean, standard deviations) on young adult NH subjects for the Dutch SSQ (Demeester et al. 2012), the English SSQ (Banh et al. 2012), and the French SSQ (Moulin et al. 2015).
- Factor analyses of the English SSQ (Akeroyd et al. 2014) and French SSQ (Moulin et al. 2015).

One of the recommended methods suggested by Widaman et al. (2011) for building short-forms is to identify a subset of items in the full-length questionnaire, which maintains the factorial integrity of the main scale. For the SSQ, factor analyses have shown that the main factors correspond to the three main subscales, Speech, Spatial, and Qualities (Akeroyd et al. 2014; Moulin et al. 2015). The recommended minimum number of items per subscale ranges between three and five (Loewenthal 2001). We opted to create a short-form with five items per subscale (15 items in total), thus keeping a possibility of further scaling down the questionnaire to three or four items per subscale. To distinguish this new short-form from other SSQ short-forms, namely, Noble et al. (2013) SSQ12 and Kiessling et al. (2011) SSQ15, we hereafter refer to it as the "15-item SSQ" (15iSSQ).

The selection of items for inclusion in the short-form used the three categories of "item quality" proposed by Stanton et al. (2002): (1) internal items qualities, which refer to scale properties such as inter-item and total-to-item correlations, internal consistency measures (e.g., Cronbach's α), factor analysis results, and subscale internal structure; (2) external item qualities, which refer to the way the items interact with external factors (construct validity), such as correlations between scores and hearing loss, ear asymmetry, age, or number of years of education. (3) Judgmental item qualities, which refers to the subjective evaluation by the scale user, of the relevance and adaptation to the patient's level of understanding. Indeed, some items of the SSQ are quite long to read and can be difficult to understand, as shown by readability scores, that range from easy to extremely difficult for some items of the quality subscale (Moulin et al. 2015). Because these indices can yield contradictory information regarding the "quality" of an item, item selection often reflects a compromise between different evaluation criteria. To deal with this, Stanton et al. proposed to code the quality indices for each item and to sort the items by their level of quality, starting from external items quality such as item-level validity, followed by judgmental quality and face validity, and finally, internal consistency item qualities. As we wanted to eliminate, first, the most unreliable items, we adapted this approach by devising a system of "penalty points," awarded for each item according to the extent of its departure from several numeric criteria related to the quality categories. The scale for "penalty points" was 0, 1, and 2; a penalty of 2 on one criterion meant that the corresponding item was eliminated. The details of the penalty points awarded to each item are provided in the Table ST1 in Supplemental Digital Content 1, http://links.lww.com/EANDH/A480. In a first phase, the following criteria were used:

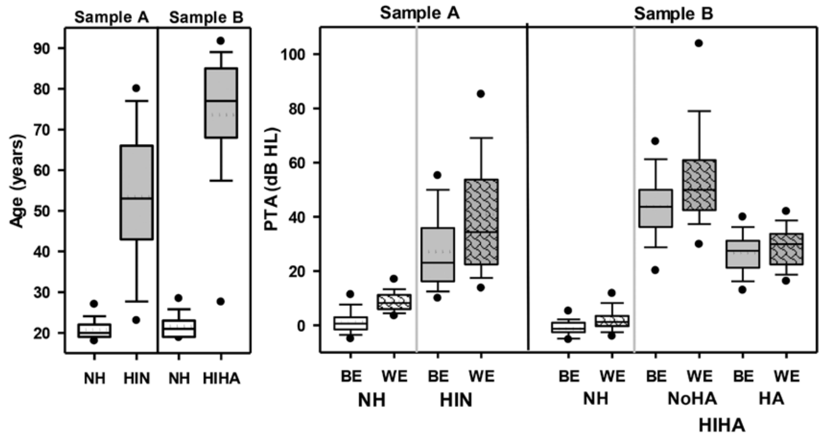

Figure 1. Statistical characteristics of the different subject samples. Left, Age. Middle and right, Best-ear (BE) and worst-ear (WE) pure-tone threshold average (PTA). For sample B, all hearing-impaired (HI) subjects were hearing aid wearers (HA), and PTAs measured with HAs and without HAs ("NoHA") are reported. HIN indicates hearing-impaired, non hearing aid wearers; NH, normal hearing.

1. Elimination of items with the greatest percentage of missing answers and non applicable responses (missing >10% in English, French, and German SSQ). Indeed, the number of missing answers (subjects who ticked the "non applicable" box) can be quite high for some items, which happen to be the same items across the different data samples (Fig. 1 showing comparison of SSQ missing answers in Moulin et al. 2015). For shorter forms, this can be a problem, if some items are systematically missing (Putnam & Rothbart 2009). Therefore, the rate of missing items was one of the criteria of choice for reducing the number of items for the short-form.
2. Elimination of items with less than 20% of variance explained by a model involving hearing loss, ear asymmetry, age, gender, and number of years of education (based on an analysis of French SSQ data from Moulin et al. 2016a). This allowed us to exclude individual items that showed low correlations with hearing loss and ear asymmetry. Hence, this criterion was used to reinforce the external validity of the scale, to ensure good correlation with external factors such as hearing loss and hearing asymmetry.
3. Elimination of all items with the lowest communalities (<50%) in factor analyses of the French SSQ and the English SSQ.
4. Elimination of items having a main factor load lower than 0.60 and/or cross-loadings larger than 0.20 (based on factor analyses of the French SSQ and the English SSQ). The 1.3 and 1.4 criteria allow to ensure the maintenance of the internal structure of the SSQ, with three well-defined subscales, in the short-form.
5. Elimination of items whose cutoff scores (defined as the mean − 2 SD), in young NH subjects, were below 3.5. This was based on 98 subjects for the French SSQ, 103 subjects for the Dutch SSQ (Demeester et al. 2012), and 48 subjects for the English SSQ (Banh et al. 2012). As the mean (and SD) calculated across items correlates significantly between languages and data samples (Moulin et al. 2015), this criterion allowed to eliminate items with too low a value (defined as a value <3.5) in NH subjects. Indeed, the SSQ items that are low in NH subjects tend to decrease the contrast between NH and HI subjects.

This first selection phase left a total of 23 items remaining, with 8 from Speech, 10 from Spatial, and 5 from Qualities.

In a second phase, the criteria listed above were used more coarsely, in that the "penalty points" were summed together rather than considered separately; in addition; four other criteria were added:

1. Favoring items with scores not significantly predicted by unwanted factors, such as the number of years of education or gender. Indeed, the scores of some items of the spatial and quality subscales showed a significant correlation (albeit minor) with the number of years of education of the patients (Moulin & Richard 2016a). Those items were among the longest and most difficult to read by the patients. Therefore, eliminating preferably those items from the short-form is likely to increase its validity.
2. Favoring, for the spatial scale, items showing a strong prediction by left/right asymmetry in pure-tone thresholds and eliminating items predicted strongly by ear asymmetry for the other subscales. Indeed, as the spatial scale is strongly related to localization in space and left/right asymmetry, favoring items correlating the most with ear asymmetry is likely to reinforce the specificity of the spatial subscale.
3. Favoring, whenever possible, items already present in the SSQ12, the SSQ15, or the SSQ5.
4. Favoring items with high item-to-total correlations. This criterion allows to increase the reliability of the short-form, but as all SSQ items show a good item-to-total correlation, the expected reliability of the short-form is high, regardless of item choice. Hence, this was not our first criteria.

This second phase led to a 15-item short-form (#1.1, #1.4, #1.5, #1.6, #1.11, #2.2, #2.6, #2.7, #2.11, #2.17, #3.4, #3.5, #3.6, #3.8, #3.9) having one item in common with the SSQ5, five items in common with the SSQ12, and seven items in common with the SSQ15. As an aside, the SSQ12 and SSQ15 have five items in common, while the SSQ5 and SSQ12 have three items in common (see Table ST1 in Supplemental Digital Content 1, http://links.lww.com/EANDH/A480). Importantly, the 15-item short-form thus created conforms to the three main subscales of the SSQ.

Short-Form Validation

The short-forms were validated with two samples of subjects. Sample A contained 98 young NH subjects (18 to 27 years of age; mean age = 20.8 years; SD = 2.2 years) and 196 HI subjects (18 to 88 years of age; age = 53.4 years; SD = 17.1 years), who did not use hearing aids (HIN). For the latter, the four-frequency (0.5, 1, 2, and 4 kHz) PTA was equal to 27.2 dB HL (SD = 14.4 dB) on the best ear and 41 dB HL (SD = 24 dB) on the worse ear, on average (Fig. 1). The HIN subjects formed a subset of an original sample of 216 HI subjects, which was used in two previous publications, along with the 98 NH subjects (Moulin & Richard 2016a 2016b).

Sample B was a new sample, completely independent from the first one. It comprised 35 NH subjects 18 to 30 years of age (mean = 21.6 years, SD = 2.7 years) and 88 HI hearing aid users (HIHA) 19 to 94 years of age (mean = 73.6 years, SD = 16.2 years). HI participants without HAs (sample A), as well as with HAs (sample B), were tested so as to document the applicability of the French language SSQ to these two types of HI participants, the French language SSQ characteristics having been only described in HI patients (and NH subjects) so far. The NH subjects were recruited mostly from a pool of university undergraduates. They were free from any medical condition, history of otologic pathology, medication, and heavy noise exposure. They underwent pure-tone audiometry in half-octave steps from 125 to 8000 Hz using an Interacoustic AC40 audiometer in a sound-proof booth. The HIHA subjects were recruited from an audiology clinic. Their unaided pure-tone thresholds were measured at 250 and 500 Hz, then at half-octave steps from 500 to 8000 Hz, using an Aurical Astera. Aided thresholds were measured using warble tones at these same frequencies, through speakers. Those patients had been wearing their hearing aids for an average of 3.4 years (SD = 3.7 years) and an average of 10.4 hr per day (SD = 5.1 hr). The average PTA for the best ear was 43.9 dB HL (SD = 13.4 dB) and 53.8 dB HL (SD = 19.5 dB) for the worse ear (Fig. 1).

Subjects from sample B were administered the validated full French language version SSQ (Moulin et al. 2015) and the new short-form version at two different time points, separated by at least 2 hr and up to a few days, in randomized testing order; the short-form was presented first to half of the subjects and second to the other half of the subjects. The HIHA subjects were instructed to fill the SSQ as if they were listening through their hearing aids. For all samples, the SSQ and the short-form were completed independently by the participants, after these were instructed by an audiologist; the interview mode was not used.

The research was conducted in agreement with the World Medical Association Declaration of Helsinki and in agreement with the French law pertaining to biomedical research (agreement number A-11–385, “CPP Sud-Est IV”).

Data and Statistical Analyses

Score Calculations • PTA was computed as the average hearing threshold (in dB HL) across 0.5, 1, 2, and 4 kHz. Ear asymmetry was calculated as the difference in PTAs (in dB) between the left and right ears. SSQ scores were calculated over the full scale and over the different short-forms that were available in the literature: the SSQ5 (#1.8, #2.3 and #2.9, #3.9 and #3.14, its weighted version (SSQ5w = [(#1.8 × 0.804) + (#2.3 × 0.770) + (#2.9 × 0.676) + (#3.9 × 0.806) + (#3.14 × 0.646)]/5) (Demeester et al. 2012)), the SSQ12 (#1.1, #1.4 and #1.10 to #1.12; #2.6, #2.9 and #2.13; #3.2, #3.7, #3.9, and #3.14; (Noble et al. 2013), and the SSQ15 (#1.4, #1.5, #1.7, #1.9, #1.10; #2.5 to #2.7, #2.9, and #2.12; #3.3, #3.4) (Kiessling et al. 2011). Scores corresponding to the new 15-item short-form, 15iSSQ, (#1.1, #1.4, #1.5, #1.6, #1.11, #2.2, #2.6, #2.7, #2.11, #2.17, #3.4, #3.5, #3.6, #3.8, and #3.9) were also calculated.

Scores were computed individually for every NH and HI subject in samples A and B. For sample B, two scores were computed for the 15iSSQ: one score was obtained by scoring a subset of items from the SSQ questionnaire, hereafter referred to as the new 15-item SSQ short-form, obtained as a subset of the SSQ (15iSSQs), for “15-item SSQ subset”; and the second score was obtained by scoring the actual 15-item short-form, which was administered separately to the subjects rather than as a part of the full SSQ. All the other short-forms were obtained as a subset of the SSQ for both samples. Hence, to avoid confusion, those other short-forms will be labeled 15-item SSQ short-form, obtained as a subset of the SSQ (SSQ15s), 12-item SSQ short-form, obtained as a subset of the SSQ (SSQ12s), and 5-item SSQ short-form, obtained as a subset of the SSQ (SSQ5s).

For the SSQ15s and the 15iSSQ, both of which were designed to maintain the three-subscale structure of the SSQ (Speech, Spatial, and Qualities), three “differential” subscores were calculated by subtracting subscores for the three subscales, pairwise (Qualities − Speech; Qualities − Spatial, and Speech − Spatial).

Similarly to Demeester et al. (2012), we calculated the SSQ-disability cutoffs, defined as the mean scores obtained in NH population −2 SDs, for the total scores obtained from SSQ and all short-forms, and the Speech, Spatial, and Qualities subscales.

Internal Structure Analysis • Data from the various short-forms were analyzed with common factor analysis, using the same parameters as for the full SSQ (see Moulin et al. 2015). Briefly, these analyses were performed using the R package “Psych,” with the following parameters: correlation matrix of 15 × 15 (SSQ15i), maximum likelihood method for factors extraction, parallel analysis for number of factors extraction, and oblique (oblimin) factor rotation.

Second, to determine whether factors of the short-forms reflect the same structure as the full scale, hierarchical cluster analyses (Revelle’s cluster algorithm ICLUST (Revelle 1978, 1979)) were performed. Revelle’s method is designed specifically to visualize questionnaire scales and subscales. It relies on two indices: the α coefficient (mean split-half reliability), a measure of internal consistency, and Revelle’s β coefficient (Zinbarg et al. 2005), a measure of factorial homogeneity; specifically, Revelle’s β coefficient is the worst (i.e., the lowest) split-half reliability of a scale, and hence is lower than α. In short, the scale structure is built starting from two item clusters that are most similar to each other; an item is added to the initial two item clusters only if this addition improves the internal consistency (measured by α) and/or the factorial homogeneity (measured by β) of the cluster. The results are shown using a hierarchical tree diagram of clusters that displays the internal substructure of the scale, allowing the definition of homogeneous subscales. The tree diagram connects increasingly less similar items and/or clusters from left to right: the most similar items are combined first, and the most dissimilar items are added last. α and β coefficients are provided for each cluster, and correlations coefficients are given for each connection between clusters and/or items. An

α above 0.8 and a β above 0.7, with a difference between these two coefficients lower than 0.1, are indicative of good homogeneity and good internal consistency (Cooksey & Soutar 2006). A goodness-of-fit index was used to compare the quality of the structures. Specifically, the fits of the different models to the data were assessed using root-mean-square residuals (RMSRs). An RMSR lower than 0.05 indicates a good fit; RMSRs comprised between 0.05 and 0.08 indicate a fair fit (Cooksey & Soutar 2006; Fabrigar et al. 1999). This technique is complementary to the more classical approach of factor analysis and is less method dependent (Cooksey & Soutar 2006). The ICLUST algorithm was applied to the whole sample. All statistical analyses were performed using the "Psych" package within the R statistical package version 3.1.0 (2014-04-10).

Comparisons Between Different Short-Forms • Short-form and subscale scores were compared with each other, and with SSQ scores. Because the data of the NH participants deviated from normality (Shapiro-Wilk test, $p < 0.001$, kurtosis between 2 and 4 for some variables), these data were compared using nonparametric tests: Wilcoxon sign test, Friedman analysis of variance (F-ANOVA), and Spearman correlation coefficient. The data of the HI subjects, and those used for whole-group analyses (including both NH and HI subjects), met the normality assumptions (|skewness| <2.6; |kurtosis| <2.1). Accordingly, for these data, parametric tests were used: analysis of variance for repeated measures, paired *t* tests, and Pearson correlation coefficients. For both NH and HI groups, internal validity of each short-form was assessed using Cronbach's α coefficient. Correlations between the SSQ and the different short-forms were calculated. To correct for the hyperinflation of those correlation coefficients due to the presence of common items in the SSQ and the short-forms, correction by Girard and Christensen (2008) for overlapping error variance was used. In order to obtain reciprocal equations allowing to transform the SSQ scores to a short-form score and vice versa, taking into account measurement error of both scores, orthogonal distance regressions were performed on the total population.

Multiple regression analysis of the SSQ, the short-forms, and the subscale scores of the HIN and HIHA subjects was performed using three predictors: better ear PTA (in dB HL), ear asymmetry (in dB), and age (in years). The normality of the residuals was checked using the Shapiro-Wilk test, and the assumptions of nonmulticollinearity were checked using Durbin-Watson test and variance inflation factor (VIF) statistics. Correlation coefficients were compared using Revelle's two-tailed test for correlated coefficients (R psych package).

ROC Analysis • ROC curves were calculated for HI and NH subjects (HI being the patients and NH subjects being the controls), for the SSQ, each of the short-forms, and each of the three subscales, using Robin et al. (2011) R package, pROC. The ROC curves were compared using Venkatraman test for paired ROCs. The areas under the ROC curve (AUCs) and partial AUC across regions of high sensitivity (90 to 100%) were compared using Robin et al. bootstrap test, based on the percentile bootstrap method, using 10,000 replicates. The *Z* statistic and two-tailed *p* value associated with Robin et al. bootstrap are given. As the SSQ is designed to assess a difference across two conditions (e.g., HI versus NH, or hearing aid versus no hearing aid), the region of high sensitivity appeared to be most relevant.

RESULTS

Factor and Cluster Analysis of the Short-Forms

To check the underlying factor structure of the various short-forms and their similarity with the SSQ, common factor analysis with maximum likelihood factor extraction was performed on all of the SSQ short-forms except for the SSQ5, as the SSQ5 is for screening purpose and does not reflect the internal three subscales structure of the SSQ. Although the SSQ12 was not meant initially to reflect the three subscales structure of the SSQ, it contains items that were belonging initially to each one of the subscales of the SSQ. Analysis of its internal structure might therefore useful in confirming, or not, that it can reflect the three subscales. This was done separately for the 196 HIN and the 88 HIHA subjects.

Factor Analysis

All the results from the factor analyses are in Table 1. The Kaiser-Meyer-Olkin indexes of Sampling adequacy were above 0.84 (which corresponds to "very good") (Field et al. 2012; Fabrigar et al. 1999), with values for individual items all higher than 0.71. Bartlett tests of sphericity were all highly significant. The analysis showed systematically three factors with eigenvalues greater than or equal to one (except for the SSQ12s), with cumulative variance explained between 65% (SSQ12s) and 75% (15iSSQ) (71% for the 15iSSQs). The three-factor structure was systematically confirmed by Cattell scree test, parallel analysis, and Velicer minimum average partial (MAP) criterion. The mean communalities ranged from 0.65 to 0.75. The RMSR ranged between 0.03 and 0.04. Hence, a three-factor extraction was always chosen for the final analysis, and an oblique rotation (oblimin) was applied, as the scores of all items were intercorrelated. The three rotated factors explained each at least 21% of the variance (17% for the SSQ12s), and each factor loaded primarily on items corresponding to the main subscales of the SSQ. The three rotated factors correlated with each other, with correlation coefficients ranging from 0.41 to 0.67. The different indices showed that this three-factor solution was adequate, with a Tucker-Lewis index above 0.9 (one exception at 0.87), a comparative fit index above 0.92, and root mean square error of approximation between 0.08 and 0.13.

For the 15iSSQ, the items loadings on each factor showed clear separation between each factor, with no consistent cross-loading across the three-factor analysis. A small degree of cross-loading was obtained for #1.5 especially for the 15iSSQs on the 88 HIHA and for #2.2 but only for 15iSSQ (Fig. 2).

For the SSQ15s, the separation between the different loads is less clear: for both samples, item #3.4 showed low communality (0.56) and cross-loading on factor 3 and factor 2 and both #3.3 and #2.12 showed a minor degree of cross-loadings on factors 2 and 3. Item #1.5 showed some cross-loadings on factors 1 and 2 for the 88 HIHA (Fig. 2). Hence, in the two different populations, items #3.3, #3.4, and #2.12 did not perform as well as the others.

For the SSQ12s, the three-factor structure is not well respected: the percentage of variance explained, the communalities were the lowest of all the short-forms (Table 1). For both samples, #3.14 loads on factor 2, instead on factor 1. Items #3.7, #3.14, and #1.12 showed low communality (<0.53) and heavy cross-loading on factors 1 and 2 for #3.14 and #1.12, and on factors 1 and 3 for #3.7. Three groups can

TABLE 1. Results of Factor Analyses Performed on 15iSSQ, 15iSSQs, SSQ15s, SSQ12s Data From the HIHA and HIN Subjects

Population Analyzed	88 HIHA patients				196 HI patients		
Form Analyzed	15iSSQ	15iSSQs	SSQ15s	SSQ12s	15iSSQs	SSQ15s	SSQ12s
KMO index	0.92	0.84	0.9	0.88	0.93	0.93	0.91
Minimum KMO value for individual items	0.86	0.71	0.73	0.79	0.88	0.86	0.88
Bartell test of sphericity (df = 105, $p < 0.00001$) (χ^2)	1300	1220	1069	676	2838	2656	1729
Eigenvalues of first 3 factors (F1, F2, and F3)	8.2, 1.7, 1.4	8.2, 1.3, 1.1	8.0, 1.3, 1.0	4.0, 3.3, 0.6	8.8, 1.5, 1.0	8.6, 1.2, 1.0	6.2, 1.2, 0.8
% variance explained by the 3 factors	55, 11, 9	55, 9, 8	54, 9, 7	33, 27, 5	58, 10, 7	58, 8, 7	52, 10, 6
Cumulative variance explained (%)	75	71	69	65	75	72	68
Mean communality (SD)	0.75 (0.12)	0.71 (0.15)	0.69 (0.10)	0.65 (0.18)	0.75 (0.08)	0.71 (0.10)	0.68 (0.16)
Minimum communality	0.59	0.49	0.55	0.42	0.6	0.55	0.41
% variance explained by the 3 rotated factors	28, 24, 24	27, 23, 21	27, 21, 21	29, 20, 17	26, 25, 24	25, 24, 23	30, 20, 18
Correlation coefficient between the 3 factors	0.44 to 0.63	0.52 to 0.67	0.45 to 0.59	0.41 to 0.64	0.54 to 0.66	0.65 to 0.66	0.62 to 0.67
Items on which F1 loads, with minimum load	Spatial, 0.57	Spatial, 0.73	Spatial, 0.62	Qualities, 0.32	Speech, 0.64	Spatial, 0.65	Qualities, 0.28
Items on which F2 loads, with minimum load	Speech, 0.65	Speech, 0.52	Speech, 0.51	Speech, 0.47	Quality, 0.65	Quality, 0.56	Speech, 0.39
Items on which F3 loads, with minimum load	Quality, 0.67	Quality, 0.57	Quality, 0.43	Spatial, 0.55	Spatial, 0.73	Speech, 0.66	Spatial, 0.73
Tucker-Lewis index	0.96	0.87	0.9	0.95	0.93	0.92	0.95
CFI	0.98	0.92	0.94	0.98	0.96	0.95	0.97
RMSR	0.03	0.04	0.04	0.04	0.03	0.03	0.03
RMSEA	0.08	0.13	0.11	0.08	0.1	0.1	0.08

15iSSQ indicates 15-item SSQ; 15iSSQs, new 15-item SSQ short-form, obtained as a subset of the SSQ; CFI, comparative fit index; df, degrees of freedom; HIHA, hearing-impaired, hearing aid wearers; HIN, hearing-impaired, non hearing aid wearers; KMO, Kaiser-Meyer-Olkin; RMSR, root-mean-square residual; RMSEA, root mean square error of approximation; SSQ, Speech Spatial and Qualities of Hearing Scale; SSQ12s, 12-item SSQ short-form, obtained as a subset of the SSQ; SSQ15s, 15-item SSQ short-form, obtained as a subset of the SSQ.

be distinguished: a group of 6 items, comprising the #3.14 and the 5 speech items; a group of three items belonging to the spatial scale; and a group of three items belonging to the qualities scale (Figure SF1 in Supplemental Digital Content 2, http://links.lww.com/EANDH/A481).

Cluster Analysis

Hierarchical cluster analysis (ICLUST) corroborated the three-factor structure, with three distinct clusters corresponding the three subscales (Speech, Spatial, and Qualities), for both the 15iSSQ and the SSQ15s forms, but not for the SSQ12s.

For the 15iSSQ (Fig. 3), α and β coefficients were higher than 0.82, which is above their usual criterion values (0.8 for α and 0.7 for β), and the difference between these coefficients was small (<0.1), indicating high homogeneity and consistency within each cluster (Cooksey & Soutar 2006). The goodness-of-fit measure for this three-subscale solution was 0.98, with an RMSR <0.05, which corresponds to an excellent fit. The α coefficient above 0.91 for the three main sub-clusters showed their high reliability.

The spatial items are represented in a close-knit cluster (C8), with α = 0.94, in which subcluster C6 can be identified and represents the items pertaining to localization (locate vehicle, dog, person).

Speech items show a close association between items 1.4, 1.11, and 1.6, as a subcluster (C5), all pertaining to auditory perception in conditions of several talkers. Items 1.1 and 1.5, concerning talking with one person with one source of noise, are associated in subcluster C9.

Within the quality items, the two items about voice pitch and familiar music are associated in cluster C3, whereas the other three items, about clarity and naturalness of sounds, are associated in subcluster C11.

For the SSQ15s, the hierarchical cluster analysis showed a three-factor structure, with each factor corresponding to the expected subscale, with an acceptable fit (0.97, RMSR = 0.05), a minimum β coefficient at 0.77, and a minimum α coefficient at 0.81. For the 196 HIN subjects, the subcluster "quality" showed a lower reliability (α = 0.89, β = 0.77), probably due to the item #3.3 (music and voice as separate items), whose addition to the cluster led to a decrease of 0.1 in the β coefficient, with a difference between α and β coefficient of 0.12 for this factor, slightly higher than the expected <0.1 (Figure SF2 in Supplemental Digital Content 3, http://links.lww.com/EANDH/A482).

For the SSQ12s, the hierarchical cluster analysis showed a three-factor structure but that did not correspond to the subscales of the SSQ. The three clusters structure showed an acceptable fit (0.94, RMSR = 0.06), a minimum β coefficient at 0.74, and a minimum α coefficient at 0.81. For the 196 HIN subjects, two three-item clusters were formed, C7 and C4, and one six-item cluster (C9). C9 is composed of a subcluster of 4 items (C6) that shows good homogeneity (α = 0.92, β = 0.89), and two items (#1.12 and #3.14), the addition of which decreases both the β and α coefficients, so that the end cluster, C9, shows a difference between β and α coefficient greater than 15. This reflects a lack of homogeneity of the cluster, with items #1.12 and #3.14 being outliers (Figure SF3 in Supplemental Digital Content 4, http://links.lww.com/EANDH/A483).

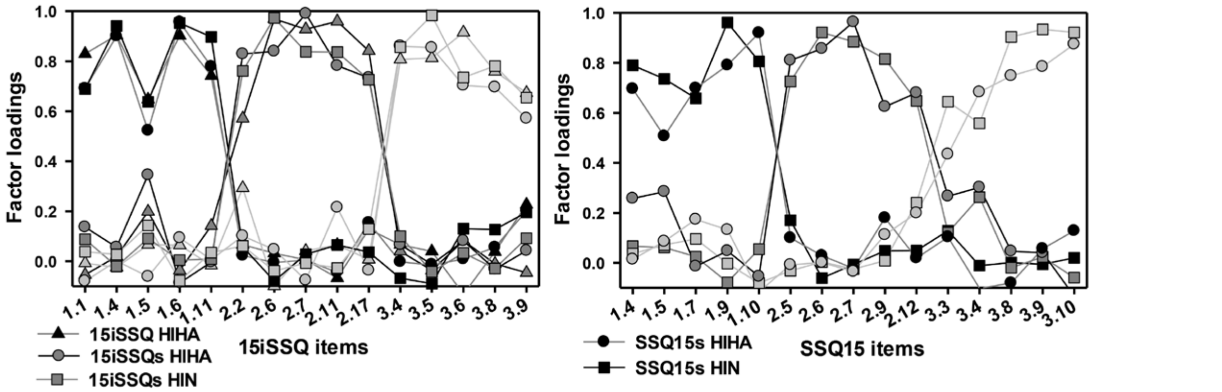

Figure 2. Factor analyses of 15iSSQ and SSQ15. Factor loadings computed using factor analysis applied on the responses to 15-item subsets from the full Speech Spatial and Qualities of Hearing Scale (SSQ) corresponding to the 15-item SSQ (15iSSQ; left) or to the 15-item SSQ short-form, obtained as a subset of the SSQ (SSQ15s; right). The three factors are indicated by different colors (black: factor 1; dark gray: factor 2; light gray: factor 3). The different subject samples are shown using different symbols, as indicated in the figure: HIHA: hearing impaired hearing aid (n = 88) and HIN: hearing impaired without hearing aids (n = 196). The new 15-item SSQ short-form, obtained as a subset of the SSQ (15iSSQs) refers to data collected separately in 88 HIHA subjects using the 15iSSQ.

Comparison Between SSQ and Short-Forms

To assess the usefulness and validity of a new tool, it is necessary to compare it to a "golden standard," here, the full-scale SSQ, and to other similar tools (here, the already existing short-forms). Indeed, it is necessary to assess how different and similar the short-forms characteristics (such as missing answer rates, scores for the global scale and the different subscales, internal reliability) are from the full scale.

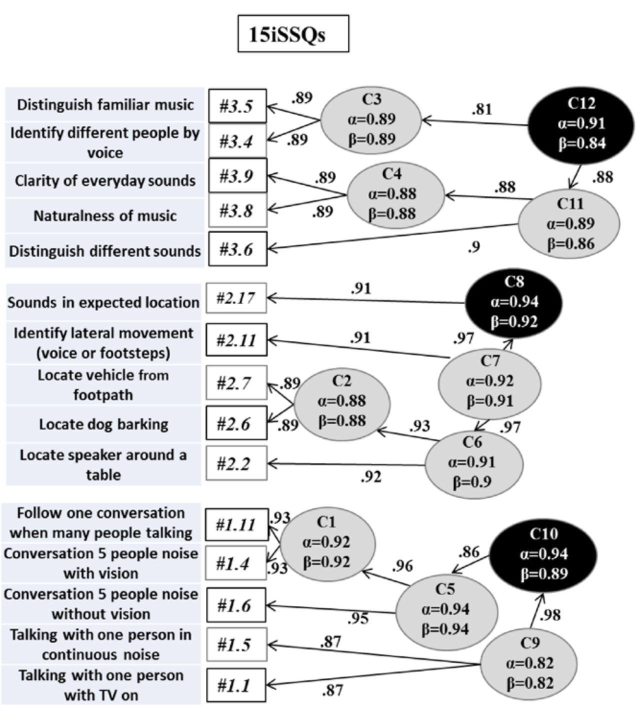

Figure 3. Tree diagram obtained using hierarchical cluster analysis (ICLUST) of the 15-item SSQ short-form, obtained as a subset of the SSQ (15iSSQs) data from 196 hearing-impaired (HI) subjects. Shortened verbal descriptions (after Banh et al. 2012) of the 15 items are listed on the left, grouped by subscale. The most similar items are combined first, and increasingly less similar clusters are represented from left to right. For each cluster, the α coefficient and Revelle's β (worst-split-half reliability) are provided. Three main clusters (in black: C8, C10, and C12), corresponding each to a main subscale, are identified.

Missing Responses • The number of missing responses was analyzed per item for the HIHA subjects (Table ST1c in Supplemental Digital Content 1, http://links.lww.com/EANDH/A480). It ranged from 0 to 14.8%, with an average of 3.1% (SD = 3.5) across the 49 items of the SSQ. Those percentages were significantly correlated with the percentages obtained in the HIN group ($r = 0.67$, $p < 0.0001$), with the same items (i.e., items 2.14 to 2.16 and 3.16) yielding the greatest number of missing responses for these two groups. The 15iSSQ gave an average of 2% (SD = 1.4) and 2% (SD = 2%) for the 15iSSQs, with a significant correlation between the two ($r = 0.72$, $p < 0.003$).

Global Scores • F-ANOVA showed highly significant differences between the different scores (full version and short-forms) for the NH subjects ($\chi^2 = 58.3$, degrees of freedom [df] = 7, $p < 0.00001$) and for the HIHA subjects ($\chi^2 = 68.4$, df = 7, $p < 0.00001$) (Fig. 4)

For NH subjects, the newly developed short-forms gave significantly greater scores than the full scale, with $W = 163$, $p < 0.02$ for the 15iSSQ, and $W = 113$, $p < 0.0006$ for the 15iSSQs. SSQ12s scores were significantly lower than the SSQ scores ($W = 563$, $p < 0.0001$). No significant differences were obtained for the SSQ15s or the SSQ5s.

For HIHA subjects, all of the short-forms except SSQ5s gave significantly lower scores than the full form, with $W = 2517$, $p < 0.03$ for the 15iSSQ, $W = 2985$, $p < 0.0001$ for the SSQ15s, and $W = 3698$, $p < 0.00001$ for the SSQ12s.

The 15iSSQ and 15iSSQs gave slightly greater cutoff scores than the SSQ in both NH samples (6.1 versus 6.0 for 35 NH), the SSQ15s (5.96) and the SSQ12s (5.31).

Subscale Scores • For NH subjects, no statistically significant difference was obtained between the full form and the different short-forms for the Speech ($\chi^2 = 11.6$, df = 5, $p<0.05$) and Spatial subscales ($\chi^2 = 6.3$, df =5, $p > 0.20$) (Fig. 4). For HIN subjects, no statistically significant difference was obtained for the Spatial subscale (F-ANOVA, $\chi^2 = 10$, df = 5, $p = 0.07$).

However, for the Speech subscale and the HIN subjects, all the different short-forms gave significantly lower scores than the full subscale (F-ANOVA $\chi^2 = 81.5$, df = 5, $p < 0.00001$), with $W = 2728$, $p < 0.0006$ for the 15iSSQ, and $W = 3606$, $p < 0.00001$ for the SSQ15s.

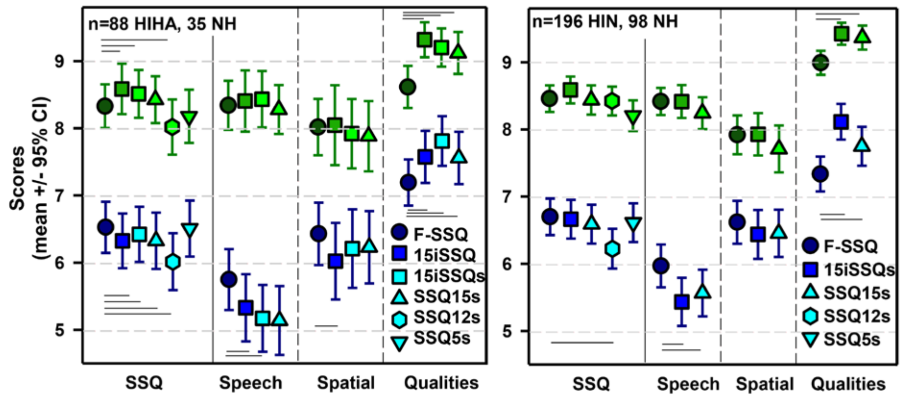

Figure 4. Scores of the different Speech Spatial and Qualities of Hearing Scale (SSQ) forms. Left, Full- and short-form SSQ scores for the 88 hearing impaired, hearing aid wearers (HIHA) subjects (blue) and 35 young normal-hearing (NH) subjects (green) composing sample B. The short-form scores were either, obtained using the 15-item SSQ (15iSSQ) or, computed by tallying scores across 5-, 12-, or 15-item subsets of the (full) SSQ corresponding to the 5-item SSQ short-form, obtained as a subset of the SSQ (SSQ5s), the 12-item SSQ short-form, obtained as a subset of the SSQ (SSQ12s), the 15-item SSQ short-form, obtained as a subset of the SSQ (SSQ15s), or the 15iSSQ; the latter subscores are referred to as new 15-item SSQ short-form, obtained as a subset of the SSQ (15iSSQs), to distinguish them from scores obtained using the 15iSSQ. Right, Full- and short-form SSQ scores obtained using subsets of the (full) SSQ in the 196 HI subjects (non hearing aid wearers: HIN, blue) and 98 young NH subjects (green) composing sample A. Error bars show 95% confidence intervals (CIs). Statistically significant differences between SSQ and the different short-forms are represented as horizontal bars (Wilcoxon tests, $p < 0.05$, see text for details).

For the Qualities subscale, highly significant differences between full form and the different short-forms were observed for NH subjects and for HIHA subjects (F-ANOVA, $\chi^2 = 68$, df = 5, $p < 0.00001$ for NH and $\chi^2 = 73$, df = 5, $p < 0.00001$ for HIHA), albeit in opposite directions: for NH subjects, all of the different short-forms yielded significantly greater Qualities subscale scores than the full form, with $W = 4$, $p < 0.00001$ for the 15iSSQ, and $W = 56$, $p < 0.0001$ for the SSQ15s; for HIHA subjects, the short-forms yielded significantly lower Qualities subscores than the full form, with $W = 1071$, $p < 0.0006$ for the 15iSSQ, and $W = 741$, $p < 0.00001$ for the SSQ15s.

Although the mean scores of the 15iSSQ Speech and Spatial subscales were not lower than the SSQ subscores, the cutoff scores of the 15iSSQ Spatial subscale were lower (4.1 versus 4.8) than the SSQ Spatial subscores, due to larger SDs. The SSQ15s gave the lowest cutoff scores (5.9 and 3.9 for Speech and Spatial, respectively). However, the Qualities cutoff scores were all greater for all short-forms (15iSSQ, SSQ15s) than for the SSQ (Table ST2 in Supplemental Digital Content 5, http://links.lww.com/EANDH/A484).

Internal Validity and Correlations Between the Different SSQ Forms • Cronbach's α was higher than 0.83 for SSQ5s, 0.93 for the SSQ12s, and 0.94 for the SSQ15s and 15iSSQ. Good consistency was also shown within each subscale, with Cronbach's α higher than 0.92 for the Speech and Spatial subscales, and comprised between 0.85 and 0.91 for the Quality subscale (Table ST3a in Supplemental Digital Content 6, http://links.lww.com/EANDH/A485). Item-to-total correlations were above 0.67 for the 196 HIN and above 0.61 for the HIHA (Table ST3b in Supplemental Digital Content 6, http://links.lww.com/EANDH/A485).

The different short-forms correlated very highly with the SSQ, with the lowest correlation coefficients obtained for the SSQ5s ($r = 0.88$, corrected Pearson) and coefficients between 0.94 and 0.95 for the other short-forms (Fig. 5). The scores for the three subscales correlated highly with the corresponding SSQ subscale scores (r above 0.91 for the Speech and Spatial subscales, and r between 0.78 and 0.85 for the Quality subscale) (Table ST4a in Supplemental Digital Content 7, http://links.lww.com/EANDH/A486). Similar results were obtained for the non-aided population (Table ST4b in Supplemental Digital Content 7, http://links.lww.com/EANDH/A486).

As the relationship between the SSQ and the SSQ12s (and to a lesser degree the SSQ15s, see Moulin & Richard 2016a) follows a power function rather than a linear one (Noble et al. 2013), we fitted the data for each of the short-forms tested using

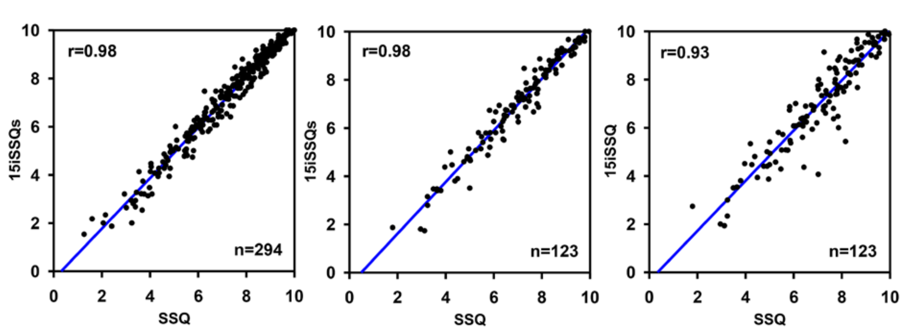

Figure 5. Scatterplots of the scores obtained for the 15-item Speech Spatial and Qualities of Hearing Scale (15iSSQ) as a function of the full SSQ score, for the 15iSSQ taken as (1) a subset of the SSQ in sample A data (left); (2) a subset of the SSQ in sample B data (middle); (3) an independent administration of the 15iSSQ in the same population as in (2) (right). Linear regressions are represented by blue lines and Pearson correlation coefficient (r) is reported for each figure.

power functions, according to Noble's formulae (with $b = 1.25$ and $c = 0.80$):

SSQ Short-forms = 10 × (SSQ/10)^{b}

SSQ = 10 × (SSQ Short-forms/10)^{c}

Nonlinear relationships between scores for the SSQ and the SSQ12s, on the one hand, and the SSQ and the SSQ15s, on the other hand, were confirmed (b =1.215 [1.180, 1.250] and b = 1.069 [1.041, 1.097], respectively). For the new short-form, the relationship was linear, as shown by a 95% confidence interval for b overlapping 1 (Table ST5 in Supplemental Digital Content 8, http://links.lww.com/EANDH/A487). Similar analyses performed on the larger sample of non hearing aid wearers (NH and HIN) yielded similar results, with a linear relationship between scores of the SSQ and of the new short-form.

Reciprocal equations obtained by orthogonal distance regression between SSQ and the new short-form, using all subjects (i.e., 417), are as follows:

15iSSQs = 0.931 × SSQ + 0.492

SSQ = 1.074 × 15iSSQs − 0.529

Construct Validity

Influence of Hearing Loss • The scores of all the different forms decreased highly significantly with increasing hearing loss in the better ear, with a slope ranging from 0.56 (SSQ5s) to 0.81 (15iSSQ) (0.65 for the SSQ) scale points per 10 dB HL. These slopes did not differ significantly from each other. However, the different short-forms differed from each other in the strength of the correlation between score and hearing loss, with a significantly higher correlation coefficient for the newly developed short-form than for the full form ($r = 0.51$ versus $r = 0.58$, $t = 2.5$, $p < 0.02$), and a significantly lower correlation for the SSQ5s than for the full form ($r = 0.41$, $t = 3.6$, $p < 0.0005$).

When analyzing subscale scores, correlations between scores and hearing loss were not found to differ significantly across the different SSQ forms (full versus short).

Linear regressions performed on the scores of the NH and HIHA subjects showed a decrease in short-form scores with hearing loss, with a greater slope for the short-forms in general, except for the SSQ5s.

SSQ = −0.065 × HL + 8.23 ($r = 0.51$, 26% variance explained)

15iSSQ = −0.081 × HL + 8.47 ($r = 0.58$, 34% variance explained)

All the details of the relationships between scores (for all the different forms and their subscales) and hearing loss are in Table ST6 in Supplemental Digital Content 9, http://links.lww.com/EANDH/A488.

Influence of Patients' Characteristics • This analysis sought to investigate predictors of the full- and short-form scores. To this aim, multiple regression analyses were performed on the scores for each form, using the following explanatory variables: gender, age, better-ear PTA, and (PTA) asymmetry. The latter two audiometric variables were measured with, and without, hearing aids, and both measurements were entered as potential predictors in the model. For these analyses, the data of the 88 HIHA subjects were used. Gender and age were never found to be statistically significant predictors. The best models included hearing thresholds measured with hearing aids, and PTA asymmetry measured without hearing aids. Therefore, multiple regression results are presented using the better-ear PTA measured with hearing aids, and PTA asymmetry measured without hearing aids. The size of the effect (r^2) and the β coefficients that allow the comparison of the relative influence of each statistically significant predictor on the SSQ scores are summarized in Table ST7 in Supplemental Digital Content 10, http://links.lww.com/EANDH/A489. The different short-forms yielded results similar to the full SSQ, with 32 to 39% of variance explained by the two predictors: better-ear PTA (β ranging from −0.28 to −0.35) and PTA asymmetry (β ranging from −0.24 to −0.19) for the seven different SSQ forms (SSQ5s, SSQ5ws, (SSQ5s with weighted items) SSQ12s, SSQ15s, l5iSSQ, l5iSSQs, SSQ). Ear asymmetry was not a significant predictor of SSQ12 scores, but was the only significant predictor of SSQ5 scores.

The three subscales of the short-forms gave similar results to the three subscales of the SSQ: better-ear PTA and asymmetry were significant predictors. For all forms, the greatest dependencies were between the asymmetry predictor and scores of the Spatial subscale and between better-ear PTA and scores of the Quality subscale. The differential scores between the main subscales (Qualities − Spatial; Qualities − Speech) did not show any significant dependency on the two predictors (Table ST7b in Supplemental Digital Content 10, http://links.lww.com/EANDH/A489).

Similar analyses were performed on the 196 non hearing aid wearers. Results showed that the different short-forms yielded results similar to the SSQ, with 37 to 38% of variance explained by the two predictors: Better Ear PTA (β from −0.54 to −0.58) and ear asymmetry (β = −0.29 to −0.32) for the five short-forms. The three subscales of the SSQ15s and 15iSSQ gave results similar to the three subscales of the SSQ, with the better-ear PTA and PTA asymmetry as significant predictors. The relative importance of PTA asymmetry for the Spatial subscale was higher for the 15iSSQs, with an even greater β coefficient (−0.47) than for hearing loss (−0.44). The differential scores (Qualities − Spatial) had PTA asymmetry as the main predictor, with 21% of variance explained for the 15iSSQs, which is substantially greater than the same score using the SSQ (11%) or the SSQ15s (10%) (Table ST7b in Supplemental Digital Content 10, http://links.lww.com/EANDH/A489).

ROC Analysis • For the SSQ, the AUC was equal to 77.3 for the HIN group and to 80.1 for the HIHA group. For the HIHA group, significantly larger AUCs (Fig. 6) were obtained for the 15iSSQ (AUC = 84) and the 15iSSQs (AUC = 83) than for the SSQ ($Z = 1.75$, $p < 0.08$, and $Z = 2.02$, $p < 0.05$ for the 15iSSQ and 15iSSQs, respectively). By contrast, AUCs for the SSQ12s and the SSQ5s were not significantly larger than the SSQ AUC. For the HIN group, differences in AUC between the SSQ and the different short-forms did not reach statistical significance (Table ST8a, b in Supplemental Digital Content 11, http://links.lww.com/EANDH/A490).

Similar analyses were performed for the three subscales (Fig. 6). For the Speech subscale, the different short-forms all showed a larger AUC (>86.4) than the full-form AUC (84.2), with significant differences for the 15iSSQs ($Z = 2.9$, $p < 0.004$ for the HIHA group; $Z = 2.4$, $p < 0.02$ for the HIN group). For the Spatial subscale, no significant difference between the different short-forms and the full form were obtained for the HIN group; however, for the HIHA group, the short-forms all produced a smaller AUC (71.9 for the 15iSSQ and 69.3 for the SSQ15s) than the SSQ (72.1), with a significant difference for the 15iSSQs ($Z = −2.2$, $p < 0.03$). For the Qualities subscale, the AUC for the 15iSSQs ($Z = −4.4$, $p < 0.0001$) was significantly

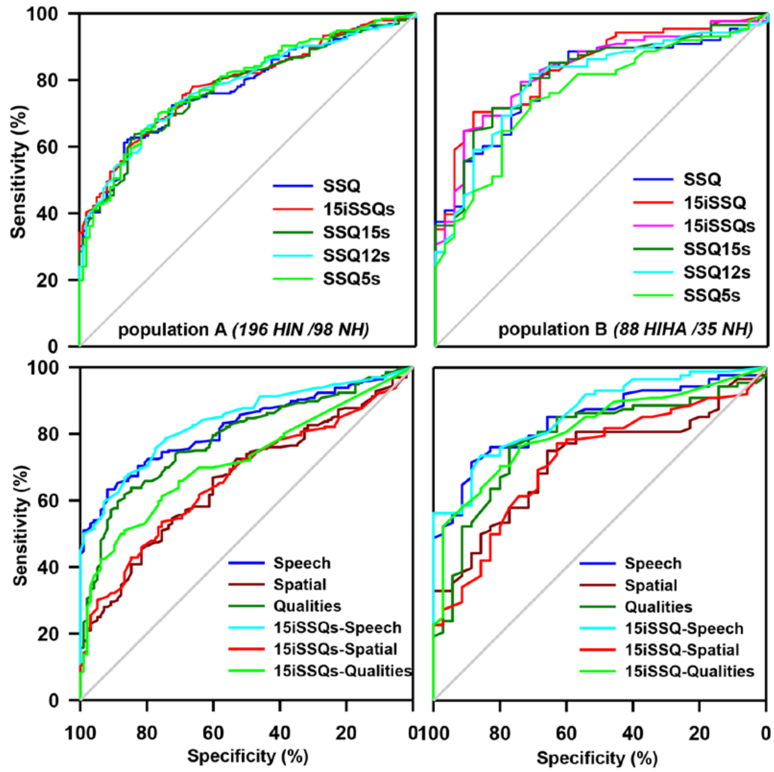

Figure 6. Receiver-operating characteristics (ROCs) for each subscale (Speech, Spatial, Qualities). Left, ROCs computed based on sample A data. Right, ROCs computed based on sample B data. Top, ROCs computed on scores for the (full) Speech Spatial and Qualities of Hearing Scale (SSQ) and for various SSQ short-forms. Bottom, ROC curves computed on subscale scores for either the (full) SSQ or the 15-item SSQ (15iSSQ).

smaller than the AUC for the full form for the HIN group, but not for HIHA group. (Tables ST8a and ST8b in Supplemental Digital Content 11, http://links.lww.com/EANDH/A490).

AUCs for the spatial subscale were systematically smaller than AUCs for the other subscales, while the AUCs for the Speech subscale were systematically larger than AUCs for the other subscales: this pattern was seen for the SSQ, as well as for the different short-forms.

A similar analysis was performed for the high-sensitivity region (sensitivity above 90%). For the HIHA group, the newly developed short-forms showed a significantly and systematically larger partial AUC than the SSQ ($Z = 2.5$, $p < 0.02$ for the 15iSSQ, and $Z = 2.4$, $p < 0.02$ for the 15iSSQs). No significant differences were obtained between the SSQ and the SSQ15s, the SSQ12s, or the SSQ5s. No significant differences were obtained for the HIN group either.

DISCUSSION

The main goal of this study was the creation of a new SSQ short-form using a rigorous data-driven approach, in agreement with the criteria suggested by Stanton et al. (2002) and the factorial integrity check suggested by Widaman et al. (2011) and Smith et al. (2000). The second goal was a detailed analysis of the SSQ (and its short-forms) using types of analysis not yet reported (for instance, ROC analysis, and internal structure of the short-forms), and involving two different samples (HIN and HIHA patients). By using data from multiple studies, we tried to avoid the caveat of building a short-form specific to a single data set (Widaman et al. 2011). By using various types of criteria for selecting items, we tried to overcome the caveats of a single-criterion approach. For instance, by using only internal consistency criteria or by over-using such criteria, one can end up with too narrowly focused a short-form, which explores only some aspects of the full form (Stanton et al. 2002). Following Stanton et al., we considered three different categories of criteria: "judgmental" (such as the percentage of missing answers, the belonging of the items to another short-form, or the readability), "external qualities" (percentage of variance explained by the five main predictors), and "internal qualities" (communalities and factor loadings of factor analysis obtained in two different studies; normative data obtained in NH subjects; inter-item and item-to-total correlations). Last, we validated the new 15iSSQ using two independent and different samples of participants, including both NH and HIHA, hence avoiding the last two of the nine "sins" of short-form building described by Smith et al.: not administering the short and full forms independently, and not using independent samples for validation.

Construct Validity

The internal structure of the short-forms was cross-checked using factor and cluster analyses. The results revealed three clusters for the SSQ15s and the 15iSSQ, each corresponding to a main subscale of the SSQ (Speech, Spatial, and Qualities). However, internal structure of the SSQ12s did not reflect the three subscales of the SSQ, with items belonging to another subscale than expected, and two items showing cross-loading between two factors. This was confirmed by cluster analysis, with one cluster of six items, itself composed of one cluster of four items plus two "outlier items," and two other clusters of three items. This was expected as the SSQ12 was not built to reflect the three subscales structure of the SSQ, but to reflect the SSQ as a whole (Noble et al. 2013) and, in particular, the 10 pragmatic subscales of the SSQ (Gatehouse & Akeroyd 2006). The 10 pragmatic subscales group items by meaning, and the SSQ12 items belong to 9 out of 10 of the pragmatic subscales (Table ST1 in Supplemental Digital Content 1, http://links.lww.com/EANDH/A480). This is why we did not calculate the scores of the SSQ12s per main subscale, as it was not relevant. SSQ12s results per main subscales are nevertheless presented in Table ST2 in Supplemental Digital Content 5, http://links.lww.com/EANDH/A484, Table ST6 in Supplemental Digital Content 9, http://links.lww.com/EANDH/A488, and Table ST7 in Supplemental Digital Content 10, http://links.lww.com/EANDH/A489.

Factor and hierarchical cluster analyses confirmed that the 15iSSQ has the same three-factor structure as the full SSQ, with no consistent cross-loading. The results further indicate that the new 15iSSQ is superior to the other existing 15-item SSQ short-form, the SSQ15s (Kiessling et al. 2011). In particular, from a construct validity perspective, factor and hierarchical cluster analyses indicated less cross-loading and better fits to the three-factor structure of the original SSQ, for the 15iSSQ (and the 15iSSQs) than for the SSQ15s.

External Validity

Regarding external validity, consistent with the full SSQ, scores for the 15iSSQ decreased with increasing hearing loss, with a higher slope for the 15iSSQ (0.81 points per 10 dB HL, 0.75 points per 10 dB HL for the 15iSSQs) than for the full SSQ (0.65 points per 10 dB HL). This outcome can be explained by one of our criteria for item selection: the presence of relatively high scores and lower variability in NH subjects. Although SSQ scores have been initially interpreted as if young NH subjects should have perfect scores (e.g., House et al. 2010), in reality, actual SSQ scores for NH are often lower than 10 (Demeester et al. 2012; Moulin et al. 2015). Indeed, in the present study, the average score of NH respondents for the Spatial subscale was lower than 8 (Fig. 4). The 15iSSQ scores were found to be significantly greater than the full-SSQ scores for both NH samples and significantly smaller for the HIN and HIHA subjects, thus reinforcing the contrast between NH and HI subjects. In addition, the expected correlation between PTA and score was significantly higher for the 15iSSQ (21% of variance explained) than for the SSQ (11% of variance explained). Unfortunately, our data sample is not sufficient to test for more subtle differences in correlations. Indeed, tests of the differences in dependent correlations require large data samples: for instance, a difference in correlation coefficient of 0.05 (0.50 versus 0.55) would require 298 subjects (with a power of 90% and a level of significance of 0.05) (Faul et al. 2009). Nevertheless, ROC analysis confirmed this greater sensitivity of the 15iSSQ (and 15iSSQs) to hearing impairment: for both HIHA and HIN groups, the AUC, a measure of the difference between the scores of the NH and HI groups, was significantly larger for the 15iSSQ (and 15iSSQs) than for the SSQ, whereas AUCs obtained using the other short-forms were not (SSQ5s, SSQ12s, and SSQ15s).

One of the caveats of reducing the number of items in a questionnaire is the potential increase of intersubject variability, the scores being less “smoothed out” by an averaging procedure across a small number of items than across a greater number of items. This has potential detrimental consequences on the disability SSQ cutoff scores. This is not the case here, as for the total score, the cutoff scores obtained with the new short-form (both 15iSSQ and 15iSSQs) were consistently larger than the scores obtained with (1) the SSQ and (2) all the other short-forms.

A look at the three main subscale scores gives a more complex picture: the cutoff scores for the Speech and Spatial subscales tend to be lower for all short-forms than for the SSQ. For the qualities subscale, all short-forms gave greater cutoff scores than the SSQ qualities. Indeed, the 15iSSQ and 15iSSQs (and to a lesser degree, the SSQ15s) gave larger scores for the Qualities subscale in all samples, especially in NH. For the Spatial and Speech subscales, 15iSSQ scores were similar to SSQ scores for both NH samples, but were significantly lower than SSQ scores for HIN and HIHA. As a result, the contrast between Qualities and Speech subscores and, to a lesser degree, between Speech and Spatial subscores, was higher for HI with the 15iSSQ than with the SSQ. The lower Spatial subscores for the HI participants may be attributed to our choice of favoring items that were more sensitive to ear asymmetry. This can be observed in the measured dependency of the Spatial scores to best ear PTA and ear asymmetry, where both predictors contributed almost equally to the scores of the 15iSSQs; by contrast, for the SSQ15s and the SSQ, the contribution of the best ear PTA was always substantially greater than the contribution of ear asymmetry. The statistical relationship between PTA asymmetry and the differential subscore, Qualities – Spatial, was substantially stronger for the 15iSSQs (21% of variance explained) than for the SSQ15s (10% of variance explained) or the SSQ (11% of variance explained). Hence, although our selection of items for the new short-form was guided primarily by the imperative to maintain the three main subscales of the SSQ, two advantageous by-products of this selection are a better contrast between NH and HI, and a Spatial subscale that is more sensitive to ear asymmetry. This greater sensitivity to ear asymmetry can be highly advantageous in the analysis of self-reported hearing disabilities linked to ear asymmetry (Vannson et al. 2015), unilateral hearing loss (Olsen et al. 2012; Dwyer et al. 2014; Douglas et al. 2007), and the evaluation of the benefits of hearing rehabilitation strategies of those asymmetrical losses (Pai et al. 2012; Dumper et al. 2009).

SSQ/Short-Form Relationships and Differences

It is worth noting that we found a linear relationship between scores obtained with the new short-form (for both 15iSSQ and 15iSSQs) and scores for the full SSQ. This differs from the SSQ12s for which the relationship is clearly nonlinear (Moulin & Richard 2016a; Noble et al. 2013) and, to a lesser degree, from the SSQ15s, and makes it straightforward to infer scores for the full SSQ based on the new short-form scores, or vice versa.

One limitation of this study stems from the fact that test–retest reliability of the new short-form was not verified on a large data sample. To the best of our knowledge, this shortcoming applies also to the other SSQ short-forms. However, it appears highly unlikely that the 15iSSQ is any less reliable than its existing 15- and 12-item counterparts, the SSQ15s and the SSQ12s. First, the items in this new short-form come from the same set of items that was used to create the SSQ15s and the

SSQ12s, namely the complete set of SSQ questions, which has shown a high degree of correlation between several administration modes and at different time points (Singh & Kathleen Pichora-Fuller 2010). Second, we obtained very good correlations between the 15iSSQs and the 15iSSQ (r=0.93) and no statistically significant differences between 15iSSQs and 15iSSQ in the same samples (NH and HIHA), whether the full score or scores per subscales were considered. Although the delay between the two administrations was too low (less than a day for about 40% of them and less than a week for most of them) to really assess long-term reproducibility, and responses to the subset version (15iSSQs) were, to some degree, influenced by the other questions of the full SSQ, it does give a clear indication of the good test/retest reproducibility of the 15iSSQ.

Even if it were merely equivalent to the full SSQ for differentiating between NH and HI subjects, the 15iSSQ could be advantageously used for this purpose in research studies or clinical work, given its shorter length. In fact, the results showed that, in some respects, the short-form affords better discrimination between NH and HI subjects than the full SSQ, and a better sensitivity to ear asymmetry thanks to the Spatial subscale. One qualification to this conclusion stems from the finding of a statistically significant smaller AUCs for the 15iSSQ than for the SSQ, for one specific combinations of HI subgroups and subscales: the Qualities subscale with the HIN group. For this reason, we cannot recommend using the 15iSSQ in lieu of the full SSQ in all situations, especially those in which longer length of the latter is not a major limitation.

Overall Differences Between Short-Forms

The differences obtained between the different short-forms appear small overall, especially the differences between the 15iSSQs and the SSQ15s. However, we need to take into account the fact that all the short-forms have been compared as subsets of a common SSQ questionnaire, hence a great proportion of data are identical across several short-forms. In particular, the 15iSSQs and the SSQ15s share 7 items out of 15, that is, almost 50% of the data are identical between the two. Nevertheless, importantly, the 15iSSQs outperform the SSQ15s in having greater cutoff values for both NH samples, for both the main scores and the three main subscales, as well as a greater dependency of its spatial subscale on ear asymmetry, a greater AUC for main scores and speech subscales, and a more defined internal structure. In addition, the independently applied 15iSSQ exhibits better characteristics than the 15iSSQs.

Although the sample sizes in this study and the redundancy of items across different short-forms contributed to limit the magnitude of any statistical differences between the 15iSSQ and the other short-forms, we found that the 15iSSQ outperformed the other short-forms in three main aspects:

- An internal structure in three clearly defined subscales that are the same as the three subscales of the SSQ. This is not the case with the SSQ12s. The SSQ15s three subscales are less clearly defined, with some cross-loading on two factors in the factorial analysis.
- A significantly greater dependency of the 15iSSQ on hearing impairment, such as hearing loss and ear asymmetry. This is shown by a significantly stronger correlation between 15iSSQ scores and hearing loss, the greater percentage of variance explained by hearing loss and ear asymmetry (for the spatial subscale) than the other short-forms, a significantly greater AUC. These results demonstrate a greater contrast between NH subjects and HI subjects.
- A linear relationship between SSQ and 15iSSQ, shown for the 15iSSQs as a subset of the SSQ, and for an independent administration of the 15iSSQ. Such a linear relationship is obtained with the SSQ5s as well, but both the SSQ12s/SSQ and SSQ15s/SSQ relationships are significantly nonlinear (power function). A linear relationship allows for easier interchangeability between SSQ and short-form scores.

ACKNOWLEDGMENTS

The authors wish to thank Jeremy Montagnat-Misson for technical help during the data collection.

C. M. is supported by Starkey Hearing Technologies, a private entity and manufacturer of hearing technology. S. G. is supported by Audition Conseil, a private company and group of audiology clinics. Other than through funding of these two coauthors' salaries, the sponsors for this study had no involvement in the design of the study, the data analysis, or the writing of the manuscript. This work was supported in part by the "Auvergne-Rhône-Alpes" region (research program "Effecbruit"); the "Fondation de l'Avenir" and "Visaudio" (research program ET4-738-VI4-001); the LABEX CELYA (ANR-11-LABX-0060) of Université de Lyon, France; and the LABEX CORTEX (ANR-11-LABX-0042) of Université de Lyon, within the program "Investissements d'Avenir" (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR).

The authors have no conflicts of interest to disclose.

Address for correspondence: Annie Moulin, Dynamique Cérébrale et Cognition (DYCOG), Centre de Recherche en Neurosciences de Lyon, CRNL, Inserm U1028-CNRS UMR5292, CH Le Vinatier, Bâtiment 452, 95 Bd Pinel, 69675 Bron Cedex, France. E-mail: annie.moulin@cnrs.fr

Received July 16, 2017; accepted September 8, 2018.

REFERENCES

Akeroyd, M. A., Guy, F. H., Harrison, D. L., et al. (2014). A factor analysis of the SSQ (Speech, Spatial, and Qualities of Hearing Scale). *Int J Audiol*, *53*, 101–114.

Banh, J., Singh, G., Pichora-Fuller, M. K. (2012). Age affects responses on the Speech, Spatial, and Qualities of Hearing Scale (SSQ) by adults with minimal audiometric loss. *J Am Acad Audiol*, *23*, 81–91; quiz 139.

Cooksey, R. W., & Soutar, G. N. (2006). Coefficient beta and hierarchical item clustering an analytical procedure for establishing and displaying the dimensionality and homogeneity of summated scales. *Organ Res Methods*, *9*, 78–98.

Demeester, K., Topsakal, V., Hendrickx, J. J., et al. (2012). Hearing disability measured by the speech, spatial, and qualities of hearing scale in clinically normal-hearing and hearing-impaired middle-aged persons, and disability screening by means of a reduced SSQ (the SSQ5). *Ear Hear*, *33*, 615–616.

Douglas, S. A., Yeung, P., Daudia, A., et al. (2007). Spatial hearing disability after acoustic neuroma removal. *Laryngoscope*, *117*, 1648–1651.

Dumper, J., Hodgetts, B., Liu, R., et al. (2009). Indications for bone-anchored hearing AIDS: A functional outcomes study. *J Otolaryngol Head Neck Surg*, *38*, 96–105.

Dwyer, N. Y., Firszt, J. B., Reeder, R. M. (2014). Effects of unilateral input and mode of hearing in the better ear: Self-reported performance using the speech, spatial and qualities of hearing scale. *Ear Hear*, *35*, 126–136.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., et al. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychol Methods*, *4*, 272–299.

Faul, F., Erdfelder, E., Buchner, A., et al. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behav Res Methods*, *41*, 1149–1160.

Field, A., Miles, J., Field, Z. (2012). *Discovering Statistics Using R* (1st ed.). Thousand Oaks, CA: SAGE Publications Ltd.
Gatehouse, S., & Akeroyd, M. (2006). Two-eared listening in dynamic situations: Audición con dos oídos en situaciones dinámicas. *Int J Audiol*., *45*, 120–124.
Gatehouse, S., & Noble, W. (2004). The speech, spatial and qualities of hearing scale (SSQ). *Int J Audiol*, *43*, 85–99.
Girard, T. A., & Christensen, B. K. (2008). Clarifying problems and offering solutions for correlated error when assessing the validity of selected-subtest short forms. *Psychol Assess*, *20*, 76–80.
Gonsalez, E. C. de M., & Almeida, K. (2015). Cross-cultural adaptation of the Speech, Spatial and Qualities of Hearing Scale (SSQ) to Brazilian Portuguese. *Audiol Commun Res*, *20*, 215–224.
House, J. W., Kutz, J. W. Jr, Chung, J., et al. (2010). Bone-anchored hearing aid subjective benefit for unilateral deafness. *Laryngoscope*, *120*, 601–607.
John, O. P., & Soto, C. J. (2009) The importance of being valid: Reliability and the process of construct validation. In Robins, R. W., Fraley, R. C., Krueger, R. F. (Eds), *Handbook of Research Methods in Personality Psychology* (1st ed.). New York, NY: The Guilford Press.
Kiessling, J., Grugel, L., Meister, I. G., et al. (2011). Übertragung der Fragebögen SADL, ECHO und SSQ ins Deutsche und deren Evaluation. German translations of questionnaires SADL, ECHO and SSQ and their evaluation. *Z Audiol*, 6–16.
Kim, B. J., An, Y. -H., Choi, J. -W., et al. (2017). Standardization for a Korean version of the Speech, Spatial and Qualities of Hearing Scale: Study of validity and reliability. *Korean J Otorhinolaryngol Head Neck Surg*, *60*, 279–294.
Loewenthal, K. M. (2001). *An Introduction to Psychological Tests and Scales*. Hove, United Kingdom: Psychology Press.
Moulin, A., Pauzie, A., Richard, C. (2015). Validation of a French translation of the Speech, Spatial, and Qualities of Hearing Scale (SSQ) and comparison with other language versions. *Int J Audiol*, *54*, 889–898.
Moulin, A., & Richard, C. (2016a). Sources of variability of speech, spatial, and qualities of hearing scale (SSQ) scores in normal-hearing and hearing-impaired populations. *Int J Audiol*, *55*, 101–109.
Moulin, A., & Richard, C. (2016b). Validation of a French-Language version of the Spatial Hearing Questionnaire, Cluster Analysis and comparison with the Speech, Spatial, and Qualities of Hearing Scale. *Ear Hear*, *37*, 412–423.
Noble, W., Jensen, N. S., Naylor, G., et al. (2013). A short form of the Speech, Spatial and Qualities of Hearing scale suitable for clinical use: The SSQ12. *Int J Audiol*, *52*, 409–412.
Olsen, S. Ø., Hernvig, L. H., Nielsen, L. H. (2012). Self-reported hearing performance among subjects with unilateral sensorineural hearing loss. *Audiol Med*, *10*, 83–92.
Pai, I., Kelleher, C., Nunn, T., et al. (2012). Outcome of bone-anchored hearing aids for single-sided deafness: A prospective study. *Acta Otolaryngol*, *132*, 751–755.
Putnam, S. P., & Rothbart, M. K. (2006). Development of short and very short forms of the Children's Behavior Questionnaire. *J Pers Assess*, *87*, 102–112.
Revelle, W. (1978). ICLUST: A cluster analytic approach to exploratory and confirmatory scale construction. *Behav Res Methods Instrum*, *10*, 739–742.
Revelle, W. (1979). Hierarchical cluster analysis and the internal structure of tests. *Multivariate Behav Res*, *14*, 57–74.
Robin, X., Turck, N., Hainard, A., et al. (2011). pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, *12*, 77.
Singh, G., & Kathleen Pichora-Fuller, M. (2010). Older adults' performance on the speech, spatial, and qualities of hearing scale (SSQ): Test-retest reliability and a comparison of interview and self-administration methods. *Int J Audiol*, *49*, 733–740.
Smith, G. T., McCarthy, D. M., Anderson, K. G. (2000). On the sins of short-form development. *Psychol Assess*, *12*, 102–111.
Stanton, J. M., Sinar, E. F., Balzer, W. K., et al. (2002). Issues and strategies for reducing the length of self-report scales. *Pers Psychol*, *55*, 167–194.
Vannson, N., James, C., Fraysse, B., et al. (2015). Quality of life and auditory performance in adults with asymmetric hearing loss. *Audiol Neurootol*, *20*(Suppl 1), 38–43.
Widaman, K. F., Little, T. D., Preacher, K. J., et al. (2011). On creating and using short forms of scales in secondary research. In K. H. Trzesniewski, M. B. Donnellan, R. E. Lucas, eds. *Secondary Data Analysis: An Introduction for Psychologists* (pp. 39–61). Washington, DC: American Psychological Association.
Zinbarg, R. E., Revelle, W., Yovel, I., et al. (2005). Cronbach's α, Revelle's β, and Mcdonald's ωH: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, *70*, 123–133.