CANADIAN JOURNAL OF SPEECH-LANGUAGE PATHOLOGY & AUDIOLOGY | CJSLPA

Volume 41, No. 3, 2017

REVUE CANADIENNE D'ORTHOPHONIE ET D'AUDIOLOGIE | RCOA

Volume 41, No. 3, 2017

SAC OAC

Speech-Language & Audiology Canada

Orthophonie et Audiologie Canada

Communicating care
La communication à coeur

CJSLPA REVIEWERS

Reviewers for this issue included: Anne Hill, Catherine Wiseman-Hakes, Chantal Desmarais, Christi Miller, Courtney Bird, Denyse Hayward, Elizabeth Rochon, Hayley Arnold, Kevin J. Munro, Sheila T. F. Moodie, Susanne Cook, and Vincent Martel-Sauvageau.

VISION AND MISSION OF SPEECH-LANGUAGE AND AUDIOLOGY CANADA

VISION
Ensuring all people of Canada achieve optimal communication health.

MISSION
Supporting and empowering our members and associates to maximize the communication health for all people of Canada.

INDEXING

CJSLPA is indexed by:

- CINAHL – Cumulative Index to Nursing and Allied Health Literature
- Elsevier Bibliographic Databases (SCOPUS)
- ProQuest – CSA Linguistics and Language Behavior Abstracts (LLBA)
- PsycInfo
- Thomson Gale – Academic Onefile
- EBSCO Publishing Inc. (CINAHL Plus with full text)
- Directory of Open Access Journals (DOAJ)

ISSN 1913-2018

SCOPE AND PURPOSE OF CJSLPA

SCOPE

The Canadian Journal of Speech-Language Pathology and Audiology (CJSLPA) is a peer-reviewed, online journal of clinical practice for audiologists, speech-language pathologists and researchers.

CJSLPA is an open access journal, which means that all articles are available on the internet to all users immediately upon publication. Users are allowed to read, download, copy, distribute, print, search, or link to the full texts of the articles, or use them for any other lawful purpose. CJSLPA does not charge publication or processing fees.

PURPOSE

The purpose of CJSLPA is to disseminate current knowledge pertaining to hearing, balance and vestibular function, feeding/swallowing, speech, language and social communication across the lifespan. Furthermore, CJSLPA is not restricted to a particular age or diagnostic group.

COPYRIGHT

CJSLPA is published by Speech-Language and Audiology Canada (SAC). Publications Agreement Number: # 40036109.
1000-1 Nicholas St., Ottawa, ON K1N 7B7 | 800.259.8519 | www.cjslpa.ca | www.sac-oac.ca

RÉVISEURS DE LA RCOA

Les personnes suivantes ont agi à titre de réviseurs pour ce numéro : Anne Hill, Catherine Wiseman-Hakes, Chantal Desmarais, Christi Miller, Courtney Bird, Denyse Hayward, Elizabeth Rochon, Hayley Arnold, Kevin J. Munro, Sheila T. F. Moodie, Susanne Cook et Vincent Martel-Sauvageau.

VISION ET MISSION D'ORTHOPHONIE ET AUDIOLOGIE CANADA

VISION

S'assurer que toutes les personnes au Canada accèdent à une santé de la communication optimale.

MISSION

Appuyer et habiliter nos membres et associés pour maximiser la santé de la communication de toutes les personnes au Canada.

INDEXATION

La RCOA est indexée dans :

- CINAHL – Cumulative Index to Nursing and Allied Health Literature
- Elsevier Bibliographic Databases (SCOPUS)
- ProQuest – CSA Linguistics and Language Behavior Abstracts (LLBA)
- PsycInfo
- Thomson Gale – Academic Onefile
- EBSCO Publishing Inc. (CINAHL Plus with full text)
- Directory of Open Access Journals (DOAJ)

ISSN 1913-2018

MISSION ET BUT DE LA RCOA

MISSION

La revue canadienne d'orthophonie et d'audiologie (RCOA) est une revue révisée par les pairs sur la pratique clinique, qui est disponible en ligne et qui est destinée aux audiologistes, orthophonistes et chercheurs.

La RCOA est une revue en accès libre, ce qui signifie que tous les articles sont disponibles sur Internet dès leur publication, et ce, pour tous les utilisateurs. Les utilisateurs sont autorisés à lire, télécharger, copier, distribuer, imprimer, rechercher ou fournir le lien vers le contenu intégral des articles, ou encore, à utiliser les articles à toutes autres fins légales. La RCOA ne charge aucun frais pour le traitement ou la publication des manuscrits.

BUT

Le but de la RCOA est de diffuser les connaissances actuelles relatives à l'audition, à la fonction vestibulaire et à l'équilibre, à l'alimentation/déglutition, à la parole, au langage et à la communication sociale, et ce, pour tous les âges de la vie. Les publications de la RCOA ne se limitent pas à un âge ou à un diagnostic particulier.

DROIT D'AUTEUR

La RCOA est publiée par Orthophonie et Audiologie Canada (OAC). Numéro de publication : # 40036109.
1, rue Nicholas, bureau 1000, Ottawa (Ontario) K1N 7B7 | 800.259.8519 | www.cjslpa.ca | www.oac-sac.ca

TABLE OF CONTENTS

TABLE DES MATIÈRES

Canadian Journal of Speech-Language Pathology and Audiology (CJSLPA)

Examining the Relationship Between Perceptions of a Known Person Who Stutters and Attitudes Toward Stuttering

Explorer la relation entre la perception des individus envers une personne bègue qu'ils connaissent et leurs attitudes face au bégaiement

Charles D. Hughes
Rodney M. Gabel
Scott T. Palasik

KEY WORDS

STUTTERING
FAMILIARITY
ATTITUDES
PERCEPTIONS

Abstract

The focus of this study was to examine the association between familiarity and attitudes toward stuttering. In total, 152 participants completed a survey consisting of Likert-type questions where they rated their perceptions of a known person who stutters (PWS). Questions were organized for analysis into 3 categories, which included perceptions of the quality of the relationship; how the known PWS copes with stuttering; and perceived impact of stuttering. Participants then completed a semantic differential scale related to their attitudes toward the known PWS, and were asked to complete the same scale thinking of an average PWS. Significant positive correlations were found between ratings of the quality of the relationship with the known PWS and positive ratings of their traits. Furthermore, how important the known PWS was to a participant was positively correlated with ratings of an average PWS as trustworthy and reliable. Perceptions regarding how the known PWS coped with stuttering were positively correlated with positive ratings of this person's traits. The most significant negative correlations were observed between perceptions of how stuttering impacted the known PWS and attitudes toward the known and average PWS. That is, the more participants perceived stuttering impacting the known PWS, the more negative their perceptions were of the known and average PWS. Findings provide support for encouraging the public to become familiar with individuals who stutter who demonstrate positive management with stuttering. Furthermore, this study helps clarify inconsistencies reported in the literature related to the impact of familiarity on attitudes toward stuttering.

Charles D. Hughes
Bowling Green State University,
Bowling Green, Ohio,
USA

Rodney M. Gabel
University of Toledo,
Toledo, Ohio,
USA

Scott T. Palasik
University of Akron,
Akron, Ohio,
USA

Abrégé

Cette étude vise à explorer la relation entre la familiarité des individus envers le bégaiement et leurs attitudes face à ce trouble de la parole. Au total, 152 participants ont rempli un questionnaire utilisant des échelles de Likert et leur demandant d'évaluer leurs perceptions envers une personne bègue qu'ils connaissent. Les questions ont été regroupées en trois catégories pour les analyses : la perception des individus concernant la qualité de leur relation avec la personne bègue qu'ils connaissent, la perception des individus quant à l'adaptation de la personne bègue qu'ils connaissent face au bégaiement et la perception des individus quant à l'impact du bégaiement. Les participants ont ensuite rempli une échelle sémantique différentielle portant sur leurs attitudes envers la personne bègue qu'ils connaissent. Ils ont également rempli la même échelle en pensant à une personne bègue typique. Les résultats montrent que la qualité de la relation des individus avec la personne bègue qu'ils connaissent est positivement et significativement corrélée avec une évaluation positive de leurs traits de personnalité. De plus, l'importance d'une personne bègue aux yeux des participants est positivement corrélée avec une perception que les personnes bègues typiques sont fiables et dignes de confiance. La perception des participants à propos de la façon dont la personne bègue qu'ils connaissent s'adapte au bégaiement est positivement corrélée avec une évaluation positive des traits de personnalité de cette personne. Les résultats montrent que les corrélations négatives les plus significatives portent sur la relation entre la perception des participants à propos de la façon dont le bégaiement affecte la personne bègue qu'ils connaissent et leurs attitudes envers la personne bègue qu'ils connaissent et les personnes bègues typiques. En d'autres mots, plus les participants perçoivent que le bégaiement affecte la personne bègue qu'ils connaissent, plus ils perçoivent négativement la personne bègue qu'ils connaissent et les personnes bègues typiques. Les résultats suggèrent que d'apprendre à connaitre une personne bègue qui prend en charge son bégaiement de façon positive devrait être encouragé au sein du public. Cette étude contribue également à clarifier les discordances rapportées dans la littérature à propos de l'impact de la familiarité des individus envers le bégaiement et leurs attitudes face à ce trouble de la parole.

It is well documented that various populations report negative attitudes toward stuttering (Cooper & Cooper, 1996; Crowe & Cooper, 1977; Crowe & Walton, 1981; Dorsey & Guenther, 2000; Silverman & Bongey, 1997; St. Louis, 2011; Turnbaugh, Guitar, & Hoffman, 1979; Walker, Mayo, & St. Louis, 2016; Yairi & Carrico, 1992). The impact of these attitudes on people who stutter has been highlighted by Yaruss and Quesal (2004) in their description of the International Classification of Functioning, Disability and Health model. In this application, Yaruss and Quesal describe how negative attitudes have an adverse impact on the quality of life of people who stutter. Therefore, it is important to explore variables that could potentially decrease these negative attitudes, in hopes of improving the quality of life of people who stutter.

One variable that has been discussed as a way to improve attitudes toward certain populations is familiarity. The benefit of familiarity can be explained through the contact hypothesis described by Allport (1954) where he suggests that, as a method to decrease stigmatization toward a marginalized group, individuals come into contact with an individual in the group in order to obtain a more accurate understanding of the population. Many studies have explored whether or not this contact, or familiarity, has an impact on attitudes toward people who stutter (Arnold & Li, 2016; Boyle, Blood, & Blood, 2009; Doody, Kalinowski, Armson, & Stuart, 1993; Gabel, Tellis, & Althouse, 2004; Hughes, Gabel, Irani, & Schalgheck, 2010; Klassen, 2001, 2002; Schlagheck, Gabel, & Hughes, 2009). Research to date has found that familiarity has an inconsistent impact on attitudes toward people who stutter.

Familiarity having no effect on attitudes

Some evidence suggests familiarity does not have an effect on attitudes. For instance, Doody et al. (1993) examined the perceptions of 106 individuals from rural communities in Newfoundland toward stuttering. They found that regardless of familiarity, participants viewed a person who stutters (PWS) more negatively versus a non-stuttering individual. Gabel et al. (2004) reported similar results in their investigation of 195 university students, which concluded that different levels of familiarity did not have a significant positive impact on perceptions toward people who stutter. Hughes et al. (2010) found similar results when examining how university students perceived the impact stuttering has on a person's life. In their survey, 110 of 146 participants reported knowing at least one PWS; however, familiarity with a PWS did not have a significant impact on perceptions. University students' attitudes toward stuttering were also explored by Boyle et al. (2009) in their investigation of 204 college-aged students. Boyle et al. examined whether causality, curability, and familiarity had an influence on attitudes toward stuttering and found that perceived causality was found to be a factor in affecting attitudes; however, familiarity was found to be unrelated to attitudes.

Familiarity having positive effects on attitudes

Other studies have shown familiarity can have a positive impact on attitudes toward stuttering. For instance, Klassen (2001) concluded that individuals who knew a PWS demonstrated a positive attitude toward people who stutter and proposed that this contact with stuttering could improve overall perceptions of people who stutter. In another study, Klassen (2002) utilized a semantic differential scale to examine responses from 108 individuals who knew someone who stutters. Klassen's findings revealed that individuals who knew someone who stutters demonstrated more positive attitudes toward stuttering when compared to previous studies of the general public toward stuttering. Klassen concluded that these findings provided support that familiarity with a PWS has a positive impact on attitudes toward stuttering. In addition, Schlagheck et al. (2009) investigated stereotyping of people who stutter using a mixed method design exploring the impact of several variables on attitudes toward stuttering, where familiarity was found to have a positive effect. More recently, Arnold and Li (2016) examined the relationship between beliefs about people who stutter and behavioural and affective reactions toward stuttering. A database from the Public Opinion Survey of Human Attributes – Stuttering was used, and when filtered for the purposes of their study produced 2,206 participants. Arnold and Li found that familiarity was related to how participants reacted toward people who stutter, and concluded that having the public become familiar with a PWS has implications related to improving how others react toward people who stutter.

Statement of the problem

Research exploring the relationship between familiarity and attitudes toward stuttering has produced mixed results. Despite the many studies that have examined this relationship, little is known as to the underlying reasons for the discrepancy. One possible explanation could be that previous studies may not have accounted for the complexity of knowing another PWS. For example, asking questions related to the extent to which a person is familiar with a PWS, and their perceptions of how they are managing their stuttering, may add another layer of understanding of the impact of familiarity on attitudes toward stuttering.

These factors could add nuances to familiarity that have not yet been fully explored, and could help clarify the varied findings observed in the relationship between familiarity and attitudes toward stuttering.

Purpose

The purpose of this study was to better understand the relationship between familiarity and attitudes toward stuttering for both a known and average PWS. The following research questions were used to explore whether a relationship exists between perceptions of a known PWS and attitudes toward the known and average PWS:

1) Does the quality of relationship with a known PWS relate to attitudes toward the known and average PWS?

2) Is there a relationship between the perceptions of how a known person manages their stuttering and attitudes toward the known and average PWS?

3) How do perceptions of how stuttering impacts a known PWS relate to attitudes toward a known and average PWS?

Methods

Questionnaire design and procedures

A questionnaire was developed after reviewing previous research exploring attitudes toward stuttering. In addition, many of the survey questions and procedures were used and adapted from previous studies (Klassen, 2001, 2002; Turnbaugh et al., 1979; Woods & Williams, 1976). One part of the questionnaire included the semantic differential scale, which has been utilized in many studies (e.g., Gabel et al., 2004; Klassen, 2001, 2002; Turnbaugh et al., 1979; Woods & Williams, 1976). This method was chosen due to the consistency in findings across studies exploring perceptions of stuttering. Additional items were designed specifically for this study to gather data about participants' demographic information, as well as perceptions of their experiences with people who stutter according to relationships, familiarity, and behaviours. Though the study did not engage in standardization and testing of the validity of these items, it was judged that these items would be appropriate for this study. These additional items were developed based on a review of published studies exploring similar research questions related to stuttering and the impact of a variety of factors on perceptions of people who stutter (Crowe & Cooper, 1977; Crowe & Walton, 1981; Doody et al., 1993; Gabel et al., 2004; Klassen, 2001, 2002; St. Louis, 2011).

The questionnaire was composed of three sections, with the first section consisting of demographic information. Some of the main questions in this section included level of education, occupation, age, gender, and if participants knew anyone who stuttered. The second section included survey questions related to perceptions toward a known PWS. If participants reported knowing multiple people who stutter, they were asked to complete the questions in the second section in regard to the person they knew the best, or in other words, with whom they were most familiar. This section asked participants to respond to questions related to the nature of their relationship and quality of familiarity with the known PWS, as well as perceptions of their communication ability and stuttering. Participants were asked to respond to questions related to quality of the relationship with the known PWS and perceptions of their communication and stuttering on a 5-point Likert scale, indicating their level of agreement from 1 (*Strongly Agree*) to 5 (*Strongly Disagree*). A total of 11 survey items used this Likert scale.

For the third and final section of the survey, participants responded to Woods and Williams' (1976) semantic differential scale, which consisted of a total of 25 items on a 7-point scale. Each item consisted of an adjective located in the left column with a corresponding antonym (e.g., trustworthy-untrustworthy) in the right column. To assure even distribution of positive and negative adjectives, each pair was randomly distributed so that positive and negative adjectives were randomly positioned in right and left columns of the scale. Participants rated their perceptions on the 7-point scale for all 25 adjectives. Participants first completed the scale with regard to the known PWS, and then completed another copy of the scale with respect to an average PWS. A definition of stuttering was not included in the survey. Thus, participants were required to think of what they believed an average, typical PWS was like when completing the second semantic differential scale. The word *average* was used as a way to keep in line with other studies that have examined attitudes toward stuttering, in that they have used a synonym of *typical* when referring to a PWS (Doody et al., 1993; Woods & Williams, 1976). Furthermore, the Klassen (2001, 2002) studies have incorporated similar procedures when measuring the impact of familiarity on attitudes toward stuttering.

Participants

This study was reviewed and approved by the Human Subjects Review Board at Bowling Green State University. In order to take part in the study, the following criteria were met: (1) being above the age of 18; (2) not

reporting a history of stuttering; and (3) knowing a PWS. Participants were recruited in a variety of settings, which included public establishments such as restaurants, office buildings, and college classrooms. A total of 326 survey packets were distributed, with 204 surveys returned. From these returned surveys, 21 were deemed incomplete, eight individuals were a PWS, and 23 did not know anyone who stuttered.

As a result, there were 152 participants who met the inclusion criteria. It should be noted that this study is part of a larger study examining factors that influence attitudes toward stuttering. Results from other parts of the survey can be found in a separate study with these participants using the previously described questionnaire procedures (Hughes, Gabel, & Palasik, 2011). Participants consisted of 65 males and 87 females with a mean age of 26.39 (SD = 12.16). A variety of relationships with people who stutter were reported, with 81 participants reporting having friends who stutter; 24 reporting having classmates/acquaintances who stutter; 13 choosing the *other* category; eight reporting a co-worker; five reporting a professor/teacher; three reporting a client, and one reporting a student. Finally, a variety of family members were reported, which included spouses, aunts/uncles, cousins, siblings, and parents—which, when combined, totalled 17 participants. Participants reported knowing some people who stutter for many years, with 65 participants reporting a relationship lasting between 1 and 10 years, and 49 participants reporting a relationship lasting longer than 10 years. Thirty-eight participants reported knowing a PWS for 1 year or less.

Analysis

Survey items. To organize the 11 Likert scale survey items for analysis, the first and second author discussed how similar questions could possibly be grouped to form categories. After multiple discussions, a consensus was reached regarding how to categorize questions. These categories are presented in Table 1. The categories consisted of questions that focused on the quality of the relationship, coping with stuttering, and the impact of stuttering. The quality of the relationship was chosen as a name for the category because these questions asked participants to reflect on how well they knew the person and how they viewed the relationship. Coping with stuttering was chosen as the descriptor for the second category because these questions generally focused on how the person dealt with their stuttering. The last category consisted of questions related to the perceived impact of stuttering on various aspects of the known person's life.

Table 1. Categories of Survey Questions

Category	Likert-type question (1 = strongly agree, 5 = strongly disagree)
Quality of Familiarity	I know this person well.
	I have a good relationship with this person.
	This person is important to me.
Coping with Stuttering	This person is a good communicator.
	This person is a competent speaker.
	This person stutters more frequently in some situations than others.
	This person appeared to be comfortable in discussing his/her stuttering.
Impact of Stuttering	I feel that stuttering has affected this person socially.
	I feel that stuttering has affected this person educationally.
	I feel that stuttering has affected this person occupationally.
	I feel that stuttering has not affected this person in any way.

Semantic differential scale. In preparation for data analysis, each of the 25 items on the semantic differential scale was scored such that the higher mean scores were indicative of a negative trait and a lower mean was indicative of a positive trait. This required that all items be arranged so the positive adjectives were allotted to the lower number on the 7-point scale, and participants' reports were adjusted accordingly. Pearson product-moment correlations were then conducted for each survey item with the 25 items on the semantic differential scale for the known and average PWS. For correlations that were significant for the known PWS but not for the average PWS, a Fisher z-test was used to transform the correlation statistic to a z-score to determine if these associations were significantly different from one another. A Bonferroni adjustment was completed with regard to the alpha level (.05) with the 25 semantic differential items. This correction was made due to the multiple comparisons with the 25 semantic differential scale items. The p-value was divided by the number of semantic differential items (.05 / 25), which equalled an alpha level of .002. This alpha level was used for analysis.

Results

Pearson product-moment correlations between the items related to the three categories (quality of relationships, coping with stuttering, and impact of stuttering) and the responses on the semantic differential scales were calculated. Additionally, descriptive statistics were calculated for each survey item (see Table 2) and for individual semantic differential scale items for the known and average PWS (see Table 3). Findings from the correlations are presented in relation to the known and then average PWS. Recall that lower numbers on the 7-point scale are related to more positive adjectives.

Table 2. Descriptive Statistics for Survey Items

Item	Mean (SD)	Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree
I know this person well.	2.21 (1.01)	45	47	46	11	3
I have a good relationship with this person.	2.16 (.931)	38	66	35	11	2
This person is important to me.	2.34 (.949)	33	50	55	12	2
This person appeared comfortable discussing his/her stuttering.	2.81 (.882)	13	32	83	19	5
I feel that stuttering has affected this person socially.	2.96 (1.15)	10	58	27	42	15
I feel that stuttering affected this person educationally.	3.39 (1.06)	3	32	44	48	25
I feel that stuttering has affected this person occupationally.	3.43 (1.00)	2	24	59	40	27
I feel that stuttering has not affected this person in any way.	3.44 (1.11)	9	23	36	60	24
This person is a good communicator.	2.58 (.903)	10	75	38	27	2
The person is a competent speaker.	2.74 (.919)	8	62	47	32	3
This person stutters more frequently in some situations than in others.	2.04 (.805)	36	82	28	4	2

N = 152

Table 3. Descriptive Statistics for Semantic Differential Scale Items

Semantic Differential Item	Known PWS mean (SD)	Average PWS mean (SD)
1. Sociable-unsociable	2.44 (1.45)	4.06 (1.34)
2. Trustworthy-untrustworthy	2.20 (1.44)	2.66 (1.20)
3. Passive-aggressive	4.29 (1.46)	4.76 (1.26)
4. Secure-insecure	3.54 (1.72)	5.08 (1.27)
5. Introverted-extroverted	3.66 (1.41)	4.84 (1.19)
6. Intelligent-dull	2.76 (1.44)	3.47 (1.22)
7. Withdrawn-outgoing	2.81 (1.59)	4.59 (1.23)
8. Hesitant-daring	3.46 (1.42)	4.80 (1.25)
9. Intelligent-unintelligent	2.77 (1.63)	3.01 (1.28)
10. Composed-anxious	4.09 (1.58)	4.72 (1.24)
11. Sincere-insincere	2.45 (1.28)	2.90 (1.32)
12. Likable-unlikable	2.06 (1.32)	2.64 (1.25)
13. Shy-bold	3.89 (1.63)	5.36 (1.36)
14. Calm-nervous	4.28 (1.52)	5.26 (1.26)
15. Pleasant-unpleasant	2.36 (1.30)	3.13 (1.30)
16. Reliable-unreliable	2.95 (1.76)	2.88 (1.31)
17. Employable-unemployable	2.39 (1.59)	2.88 (1.33)
18. Fearless-fearful	3.80 (1.34)	4.55 (1.08)
19. Friendly-unfriendly	2.02 (1.19)	2.99 (1.23)
20. Open-guarded	3.25 (1.66)	4.49 (1.45)
21. Competent-incompetent	2.70 (1.48)	3.26 (1.31)
22. Excited-frustrated	3.59 (1.39)	4.66 (1.28)
23. Sensitive-insensitive	2.84 (1.32)	2.86 (1.28)
24. Self conscious-self assured	4.07 (1.61)	5.11 (1.39)
25. Relaxed-tense	3.90 (1.45)	4.93 (1.26)

N = 152

Quality of relationship

Known PWS. Correlation results for quality of relationship survey items with the 25-item semantic differential scale of the known and average PWS can be observed in Table 4. Findings show a significant positive relationship between all three quality of relationship survey items and certain traits. For example, the more participants reported knowing someone who stutters, the more they perceived that person as sociable. In addition, the more participants perceived a good relationship with the known PWS, the more they viewed that person as trustworthy. Finally, the more participants viewed their relationship with the known PWS as important, the more they perceived that person as sociable, trustworthy, sincere, reliable, and relaxed.

Average PWS. Two significant positive correlations were observed between how important participants viewed the relationship with the known PWS and two semantic differential scale items. More specifically, the more importance participants assigned to their relationship with a known PWS, the more they perceived an average PWS to be reliable and trustworthy.

Correlation comparisons. Recall that, as part of the analysis, a Fisher z-test transformation was conducted for correlations that were found to be significant for the known PWS but not the average PWS on Likert scale items. Results of the Fisher z-test found significant differences between correlations for quality of relationship survey items and attitudes toward the known and average PWS. More specifically, the association with how well participants knew the known PWS was significantly stronger than the average PWS related to how social they viewed the person ($Z = 3.06$, $p = .002$). Furthermore, the association with how important participants viewed the relationship with the known PWS was significantly stronger for the known PWS compared to the average PWS, in regard to being social ($Z = 2.35$, $p = .018$) and relaxed ($Z = 2.33$, $p = .019$). That is, these associations did not transfer to people who stutter in general, but were found to be significantly stronger for the known person.

Table 4. Correlations for Semantic Differential Scale Items and Quality of Relationship for Known and Average PWS

Semantic Differential Scale Item	I know this person well.		I have a good relationship with this person.		This person is important to me.	
	Known	AVG	Known	AVG	Known	AVG
Sociable-unsociable	.275*	-.072	.241	-.003	.256*	-.011
Trustworthy-untrustworthy	.165	.076	.251*	.140	.297*	.249*
Passive-aggressive	.227	.041	.189	.114	.191	-.007
Secure-insecure	.109	-.034	.126	-.062	.113	-.105
Introverted-extroverted	.119	.039	.133	.018	.135	-.081
Intelligent-dull	.126	-.074	.178	-.010	.192	.050
Withdrawn-outgoing	.198	-.032	.183	-.028	.215	-.004
Hesitant-daring	.098	-.044	.053	-.056	.079	-.075
Intelligent-unintelligent	.110	.081	.091	.005	.124	.124
Composed-anxious	-.115	.089	-.014	.057	-.046	.053
Sincere-insincere	.115	.006	.126	.019	.263*	.053
Likeable-unlikable	.050	.034	.111	-.017	.206	.049

Shy-bold	.126	-.026	.094	-.057	.143	-.079
Calm-nervous	.060	.013	.101	-.060	.150	-.070
Pleasant-unpleasant	.128	.109	.142	-.013	.238	.060
Reliable-unreliable	.128	.099	.235	.124	.267*	.261*
Employable-unemployable	.051	.156	.041	.128	.103	.248
Fearless-fearful	.129	-.107	.107	-.085	.144	-.050
Friendly-unfriendly	.134	.135	.117	.083	.211	.134
Open-guarded	.106	.014	.128	-.021	.147	-.013
Competent-incompetent	.095	.003	.065	.068	.140	.103
Excited-frustrated	.066	-.061	.057	-.069	.101	-.045
Sensitive-insensitive	.055	.104	.081	.119	.133	.142
Self conscious-self assured	.064	.011	.027	-.019	.049	.011
Relaxed-tense	.159	-.020	.179	-.070	.261*	-.003

N = 152, *p < .002

Coping with stuttering

Known PWS. Complete results of the correlations for known and average PWS related to coping with stuttering can be found in Table 5. Significant positive correlations were found between perceptions of how the known PWS was comfortable in discussing their stuttering and ratings on semantic differential scale items for the known PWS. More specifically, the more participants perceived the known PWS as being comfortable discussing stuttering, the more they perceived the person as sociable, open, and relaxed. Many significant positive correlations were found between both questions related to viewing the known PWS as a *good* or *competent* communicator and semantic differential scale items. For example, the more participants perceived the known PWS as a good communicator, the more they rated them as being sociable, trustworthy, secure, extroverted, intelligent, outgoing, daring, sincere, bold, calm, pleasant, fearless, friendly, open, excited, self-assured, and relaxed. Furthermore, the more participants rated the known PWS as a competent speaker, the more they perceived them as sociable, trustworthy, secure, extroverted, intelligent, outgoing, daring, bold, competent, self-assured, and relaxed. Perceptions of the variability of the known person's stuttering produced significant negative correlations, revealing that the more participants perceived the person's stuttering as varying across situations, the more they perceived that person as anxious, nervous, and self-conscious.

Average PWS. Although some correlations were noted as approaching the level of significance, no statistically significant correlations were noted between any semantic differential scale items for the average PWS and survey items related to coping with stuttering. More specifically, perceptions of how the known PWS coped with stuttering were not found to be significantly related to attitudes toward the average PWS.

Correlation comparisons. Significant differences were found for correlations between the known and average PWS related to how good a communicator the known PWS was perceived. These significant differences were noted for the following traits: sociable ($Z = 3.94$, $p < .001$); intelligent ($Z = 2.89$, $p = .003$); outgoing ($Z = 2.52$, $p = .011$); daring ($Z = 2.69$, $p = .007$); intelligent, as compared to unintelligent ($Z = 2.33$, $p = .019$); bold ($Z = 2.17$, $p = .030$); calm ($Z = 2.20$, $p = .027$); pleasant ($Z = 2.94$, $p = .003$);

and relaxed (Z = 2.09, p = .036). In regard to perceptions of competence, significant correlation differences were found related to being sociable (Z = 3.27, p = .001), extroverted (Z = 2.39, p = .016), outgoing (Z = 2.68, p = .007), daring (Z = 3.59, p < .001), and bold (Z = 2.93, p = .003). For perceptions of the variability of stuttering, significant differences in correlations were found with the trait of being composed (Z = -2.10, p = .035). In regard to discussing stuttering openly, significant differences were found between known and average PWS ratings of being social (Z = 2.78, p = .005), open (Z = 2.49, p = .012), and relaxed (Z = 2.88, p = .004). In other words, these perceptions related to coping with stuttering were found to be stronger in association to a known PWS and did not relate to perceptions of people who stutter in general.

Table 5. Correlations for Semantic Differential Scale Items and Coping With Stuttering for Known and Average PWS

Semantic Differential Scale Item	This person is a good communicator.		This person is a competent speaker.		This person stutters more frequently in some situations than others.		This person appeared to be comfortable in discussing his/her stuttering.	
	Known	AVG	Known	AVG	Known	AVG	Known	AVG
Sociable-unsociable	.483*	.070	.427*	.077	-.004	-.014	.253*	-.063
Trustworthy-untrustworthy	.371*	.221	.255*	.195	.050	.007	.161	.208
Passive-aggressive	.118	.084	.181	-.062	-.219	-.181	-.070	.030
Secure-insecure	.296*	.093	.258*	.063	-.149	-.159	.073	-.046
Introverted-extroverted	.262*	.148	.268*	-.002	-.116	-.118	.113	-.042
Intelligent-dull	.470*	.173	.341*	.228	-.100	-.120	.244	.040
Withdrawn-outgoing	.374*	.101	.347*	.051	-.041	-.145	.196	.001
Hesitant-daring	.333*	.035	.353*	-.047	-.161	-.077	.166	.072
Intelligent-unintelligent	.434*	.192	.341*	.227	-.028	-.077	.163	.083
Composed-anxious	.169	.120	.243	.145	-.278*	-.042	.130	-.079
Sincere-insincere	.264*	.126	.187	.142	.175	.004	.218	.063
Likeable-unlikable	.221	.058	.073	.101	.135	.008	.203	.009
Shy-bold	.329*	.090	.299*	-.031	-.067	-.128	.156	.024
Calm-nervous	.270*	.022	.193	.043	-.284*	-.128	.140	.034
Pleasant-unpleasant	.257*	-.077	.045	.079	.069	-.018	.117	-.082
Reliable-unreliable	.079	.153	.062	.150	.169	.023	-.023	.043
Employable-unemployable	.195	.146	.135	.208	.029	-.014	.096	.037

Fearless-fearful	.257*	.118	.220	.027	-.134	.163	.090	.007
Friendly-unfriendly	.286*	.078	.163	.178	.179	-.080	.212	.052
Open-guarded	.283*	.084	.204	-.056	-.062	.074	.296*	.017
Competent-incompetent	.191	.106	.315*	.174	-.023	.078	.194	.073
Excited-frustrated	.337*	.184	.237	.148	.003	-.160	.217	.006
Sensitive-insensitive	.083	-.005	.004	-.020	-.032	.127	.230	.041
Self conscious-self assured	.350*	.154	.327*	.158	-.279*	-.170	.117	.109
Relaxed-tense	.415*	.197	.369*	.185	-.133	-.207	.312*	-.011

N = 152, *p < .002

Impact of stuttering

Known PWS. Table 6 displays the results of the correlation analysis between semantic differential scale items for the known and average PWS and impact of stuttering survey items. Many significant negative correlations were shared among the perceptions of how stuttering impacted the known person educationally, occupationally, and socially. The more participants believed the known PWS was impacted in these areas, the more likely they perceived them as unsociable, insecure, dull, withdrawn, hesitant, fearful, self-conscious, and tense. All three questions produced a number of significant negative correlations with semantic differential scale items, with social impact revealing 11, occupational revealing 11, and educational revealing 12. Significant positive correlations were observed between responses in regard to stuttering not having an effect and traits of being social, secure, outgoing, daring, bold, fearless, and self-assured. Overall, the higher participants rated that stuttering had an impact on the person socially, educationally, or occupationally, the more likely the respondents favoured the negative traits.

Average PWS. Numerous significant negative correlations were observed for this category for semantic differential scale items for the average PWS. In other words, the more participants perceived that stuttering had an impact on the known PWS, the more negative their attitudes were toward an average PWS. Ratings for participant responses in regard to the social impact of stuttering for the known person were related to responses to the traits of untrustworthy, dull, unreliable, and unfriendly. There were significant negative correlations noted for perceptions of how stuttering affected the known person educationally with untrustworthy and unintelligent. In addition, responses for stuttering affecting the known PWS occupationally were associated with ratings on the semantic differential scale item for unintelligent. No significant correlations were noted between responses of stuttering not having an effect on the known person and semantic differential items for the average PWS.

Correlation comparisons. For stuttering affecting the known person socially, significant differences were found in correlations with traits of being social (Z = -3.35, p = .0008), secure (Z = -2.27, p = .023), outgoing (Z = -2.95, p = .003), daring (Z = -4.4, p < .001), bold (Z = -3.94, p < .001), fearless (Z = -2.73, p = .006), self-assured (Z = -3.36, p < .001), and relaxed (Z = -2.15, p = .031). For affecting the person educationally, significant differences were found related to being secure (Z = -2.35 p = .018), intelligent (Z = -2.01, p = .044), outgoing (Z = -1.97, p = .048), daring (Z = -2.01, p = .044), and self-assured (Z = -2.28, p = .022). In regard to occupational affect, significant differences were found for being social (Z = -3.55, p < .001), secure (Z = -3.02, p = .002), daring (Z = -2.92, p = .003), bold (Z = -2.50, p = .012), fearless (Z = -2.79, p = .005), and self-assured (Z = -3.09, p = .002). Finally, for the question of stuttering not affecting the known person, significant differences were found in correlations for the traits of being social (Z = 2.99, p = .002), secure (Z = 2.30, p = .021), and daring (Z = 2.97, p = .003). In summary, these associations between how stuttering impacts a person's life were found to be significantly stronger for the known PWS compared to an average PWS, revealing that these ratings affected a known PWS, yet did not translate to associations of an average PWS.

Table 6. Correlations for Semantic Differential Scale Items and Impact of Stuttering for Known and Average PWS

Semantic Differential Scale Item	I feel that stuttering has affected this person socially.		I feel that stuttering has affected this person educationally.		I feel that stuttering has affected this person occupationally.		I feel that stuttering has not affected this person in any way.	
	Known	AVG	Known	AVG	Known	AVG	Known	AVG
Sociable-unsociable	-.368*	.002	-.306*	-.096	-.325*	.074	.357*	.027
Trustworthy-untrustworthy	-.183	-.270*	-.275*	-.280*	-.240	-.212	.126	.189
Passive-aggressive	-.226	-.108	-.010	-.102	-.096	.016	.199	.116
Secure-insecure	-.330*	-.080	-.355*	-.098	-.337*	-.001	.374*	.126
Introverted-extroverted	-.191	-.048	-.062	-.082	-.093	-.087	.112	.124
Intelligent-dull	-.265*	-.288*	-.424*	-.216	-.275*	-.205	.208	.209
Withdrawn-outgoing	-.400*	-.082	-.354*	-.141	-.269*	-.087	.338*	.128
Hesitant-daring	-.456*	.017	-.334*	-.114	-.366*	-.045	.401*	.080
Intelligent-unintelligent	-.225	-.247	-.417*	-.306*	-.361*	-.287*	.215	.232
Composed-anxious	-.111	-.096	-.048	-.103	-.107	-.084	.088	.089
Sincere-insincere	-.105	-.133	-.226	-.143	-.243	-.138	.012	.116
Likeable-unlikable	-.003	-.194	-.126	-.092	-.170	-.180	-.036	.077
Shy-bold	-.457*	-.037	-.245	-.070	-.288*	-.007	.322*	.124
Calm-nervous	-.283*	-.103	-.281*	-.118	-.250	.009	.230	.087
Pleasant-unpleasant	-.114	-.138	-.170	-.038	-.139	-.034	.038	-.059
Reliable-unreliable	-.043	-.258*	-.222	-.196	-.116	-.177	-.003	.118
Employable-unemployable	-.154	-.190	-.338*	-.179	-.279*	-.235	.073	.112
Fearless-fearful	-.371*	-.073	-.303*	-.146	-.304*	.009	.257*	.122
Friendly-unfriendly	-.116	-.266*	-.122	-.143	-.135	-.204	.003	.212
Open-guarded	-.113	.091	-.223	-.033	-.114	.002	.131	.025
Competent-incompetent	-.100	-.072	-.185	-.162	-.223	-.129	.033	.047
Excited-frustrated	-.267*	-.054	-.138	-.081	-.133	-.054	.122	.125
Sensitive-insensitive	-.057	.050	-.183	.050	-.169	-.005	.007	-.213
Self conscious-self assured	-.428*	-.068	-.302*	-.048	-.357*	-.016	.361*	.157
Relaxed-tense	-.313*	-.075	-.308*	-.140	-.327*	-.161	.189	.173

N = 152, *p < .002

Discussion

The present study explored the correlation between perceptions of a known PWS and attitudes toward stuttering for the known and an average PWS. These perceptions of a known PWS were in regard to the quality of relationship with the known person, how the person managed their stuttering, and the impact stuttering had on their life. Results from this study provide further clarification regarding how familiarity with a PWS can have the potential to improve attitudes toward stuttering.

The first interesting finding from this study was the importance of experiences and contact with a PWS. Positive attitudes toward the known PWS were related to how important participants viewed the relationship, how well they knew them, and whether they had a good relationship with this person. More favourable perceptions of the relationship with a known PWS were associated with high ratings of an average PWS as being trustworthy and reliable. These findings are consistent with other studies that found that familiarity had a positive effect on attitudes toward stuttering (Klassen, 2001, 2002; Schlagheck et al., 2009). A closer look at the methodology of one of the Klassen (2001) studies helps to understand this similarity in the findings. The participants in that study were individuals who were identified as having a close relationship with a PWS. The closeness, or quality, of this relationship may help to explain why familiarity had a positive effect on attitudes. Our findings also support that closeness is an important aspect of familiarity. Simply knowing a PWS may not improve attitudes toward stuttering; however, the association between familiarity and attitudes toward stuttering appears to be stronger if the known person is important to the respondent. The number and type of questions asked related to familiarity may also explain the similarity in other studies that found similar results. For example, our study asked a number of questions regarding perceptions toward a known PWS to capture the complexity of familiarity. In another study that found familiarity to have a positive effect on attitudes, Schlagheck et al. analyzed responses of 154 individuals who did not stutter using open- and closed-ended questions to describe the person they knew who stuttered. The use of open-ended questions may have allowed Schlagheck et al.'s participants the opportunity to expand on their perceptions of the person they knew. The factors of closeness with a known PWS, and asking more questions about the nature of familiarity, may help to explain the discrepancy with other studies that found familiarity to have no effect on attitudes (Boyle et al., 2009; Doody et al., 1993; Gabel et al., 2004; Hughes et al., 2010).

Perceptions of how the known PWS coped with stuttering were also a significant factor related to positive traits for the known PWS. It is important to discuss findings related to coping as it relates to stuttering severity and avoidance behaviours. It is possible that participants demonstrated wide variability related to how they defined and perceived effective coping with stuttering. Participants may have perceived the known person more positively because they were demonstrating a mild stuttering severity, which could be related to the known person demonstrating avoidance behaviours, thus providing the perception of a more fluent speaker. Participants may have been using the amount of stuttering as a way to judge whether the person was effectively communicating or managing their stuttering. This discussion point also relates to prior research that has found as stuttering frequency increases, listeners demonstrate more negative evaluations of a person's speech (Panico, Healey, Brouwer, & Susca, 2005). Avoidances related to stuttering moments is a real possibility and could have been perceived as effective coping with stuttering. Nevertheless, the more participants believed the known person positively coped with their stuttering, the more positive their attitudes were toward the known PWS.

The idea that simply decreasing stuttering moments might translate to increased perceptions of positive coping and managing stuttering brings up the topic of how people who stutter may perceive role models who stutter. Hughes, Gabel, Goberman, and Hughes (2011) discussed role models for people who stutter as part of their qualitative study of adults who stutter. The participants in this study reported that when they were younger, they wanted role models to assist them in managing their stuttering. The use of role models who are dealing with stuttering in a positive way could have implications not only for public attitudes, but also to help individuals cope effectively with stuttering. Reitzes (2006) also noted the importance of providing mentors who stutter to school-age children who stutter in his description of how an older child helped to mentor a younger child who stuttered in a school setting. Furthermore, Reitzes provided a review of the connection between mentorship and coping with stuttering. Our findings related to coping could also be applied to other perceptual studies related to speech therapy. For instance, Gabel (2006) found that individuals perceived a PWS more positively if the person was involved in speech therapy and they demonstrated a more mild stuttering severity. The participants in Gabel's study may have believed that speech therapy was improving the person's ability to cope and manage their stuttering. If people who do not stutter believe that

speech therapy improves how a PWS copes with their stuttering, this begs the question, "Do listeners believe that a mild stuttering severity is related to effective coping with stuttering?" More research is needed to better understand perceptions of how individuals cope with stuttering.

Perceptions of the impact of stuttering on the known PWS were also an important factor. Perceiving the known PWS as not being impacted by their stuttering was correlated with positive attitudes toward the known and average PWS. Similar to perceptions of coping with stuttering, exposing the public to individuals who have decreased the negative impact of stuttering on their life could improve public awareness and attitudes toward stuttering. This is where people in role model positions (e.g., professional athletes, actors, and other celebrities) can play an important part in helping the general public, along with people who stutter, improve their perceptions toward stuttering.

Correlation comparisons between the known and average PWS indicated stronger associations with certain survey and semantic differential scale items for the known compared to the average PWS. In general, knowing a PWS well, perceiving they are positively coping with stuttering, and believing their stuttering does not negatively impact their life was related to positive attitudes toward this particular person. Yet, these same perceptions did not translate to people who stutter in general. One possible explanation of this finding may be related to participants viewing the known person based upon that person's unique, individual characteristics, as well their personal experience with the known PWS, and not basing their perceptions on one characteristic of the person. For example, the personal experiences with the known PWS may have involved participants learning that they are a supportive friend, fun to be around, and a good person. These types of experiences may have contributed to the stronger correlations with the known compared to an average PWS.

Limitations

There are several important limitations to this study. First, the research design utilized a convenience sample, which impacts how the results can be generalized to a larger population. Also, this study used a quantitative design to explore familiarity and perceptions of a known PWS. Thus, participants were not provided with the opportunity to elaborate on responses due to all questions being in a closed-ended format. Qualitative or mixed methods designs could potentially provide in-depth knowledge regarding participant responses, thus adding to the richness of the topic. Finally, the correlation analysis conducted is unable to determine whether familiarity with a known PWS causes attitudes toward stuttering to improve; however, it does provide insights into the relationship that exists.

Another limitation is in respect to the decision to use the word "average" in the survey protocol. When participants were reflecting on an average PWS, they may have mentally visualized someone who was anywhere along the spectrum of stuttering severity. This same mental representation might have been applied to an average PWS. Therefore, using the word "average" may have skewed the results in that participants may have responded to questions with this mental representation related to stuttering severity. In retrospect, asking participants to provide some descriptions of how they perceived "average" might have helped control this term more. Future studies might provide a description of the stuttering severity rating to help participants mentally represent a consistent hypothetical PWS.

In addition, another limitation is that there may have been confusion regarding whether participants really knew someone who stutters. The level of familiarity with the known PWS, along with participants' knowledge about stuttering, could have influenced their responses. Furthermore, it could be suggested that participants may not really have known a PWS; rather, they may have known someone who was highly disfluent or demonstrated some other communication disorder. Again, having participants describe the person they know who stutters and some of their behaviours might clarify any confusion and address this potential limitation.

Despite these limitations, the current findings have implications for people who stutter. For example, encouraging a person who stutters to have quality interactions with others, where they get to know other people in a meaningful way and view the relationship as good and important, could possibly help to improve the attitudes of people who do not stutter toward stuttering. Furthermore, we can speculate that in the context of this meaningful relationship, others may become more familiar with how stuttering impacts them and their coping style with stuttering. With these quality relationships, people who stutter may then be able share, and others then learn, that stuttering is a piece of who they are and may not have a negative impact in areas of their life such as their occupation, educational experiences, and social life.

Future research

It is recommended that future studies explore familiarity with stuttering using mixed methods designs and qualitative approaches. The use of these methodologies may allow future participants to elaborate on their responses. It is also suggested that other researchers examine the extent to which other populations, such as employers and individuals in the helping professions, report familiarity levels with people who stutter to determine if this is a contributing factor to attitudes toward stuttering in general. Finally, the extent to which stuttering severity is factored into the question of familiarity has yet to be determined. The additional testing of these variables could provide further information to explain the complexity of knowing a PWS and its impact on attitudes toward stuttering.

Conclusions

In summary, our findings help to shed light on the complexities of familiarity and its relationship with attitudes toward stuttering. We examined other intricacies of familiarity, which involved perceptions of the quality of the relationship, impact of stuttering, and coping ability of a known PWS, and the relationship of these factors to attitudes. Our results support the idea that familiarity with a known PWS is associated with improved attitudes toward this particular person. Although more significant associations were found between familiarity and attitudes for the known PWS, familiarity was also found to be related to more favourable attitudes toward an average PWS on certain traits.

References

Allport, G.W. (1954). *The nature of prejudice*. Cambridge, MA: Addison-Wesley.

Arnold, H. S., & Li, J. (2016). Associations between beliefs about and reactions toward people who stutter. *Journal of Fluency Disorders, 47*, 27–37. doi: 10.1016/j.jfludis.2015.12.004

Boyle, M. P., Blood, G. W., & Blood, I. M. (2009). Effects of perceived causality on perceptions of persons who stutter. *Journal of Fluency Disorders, 34*, 201–218. doi: 10.1016/j.jfludis.2009.09.003

Cooper, E. B., & Cooper, C. S. (1996). Clinician attitudes towards stuttering: Two decades of change. *Journal of Fluency Disorders, 21*, 119–135. doi: 10.1016/0094-730X(96)00018-6

Crowe, T. A., & Cooper, E. B. (1977). Parental attitudes toward and knowledge of stuttering. *Journal of Communication Disorders, 10*(4), 343–357. doi: 10.1016/0021-9924(77)90031-4

Crowe, T. A., & Walton, J. H. (1981). Teacher attitudes toward stuttering. *Journal of Fluency Disorders, 6*(2), 163–174. doi: 10.1016/0094-730X(81)90013-9

Doody, I., Kalinowski, J., Armson, J., & Stuart, A. (1993). Stereotypes of stutterers and nonstutterers in three rural communities in Newfoundland. *Journal of Fluency Disorders, 18*(4), 363–373. doi: 10.1016/0094-730X(93)90015-V

Dorsey, M., & Guenther, R. K. (2000). Attitudes of professors and students toward college students who stutter. *Journal of Fluency Disorders, 25*, 77–83. doi: 10.1016/S0094-730X(99)00026-1

Gabel, R., Tellis, G., & Althouse, M. T. (2004). Perceptions of people who stutter: Effects of familiarity. In A. Packman, A. Meltzer, & H. F. M. Peters (Eds.), 4th world congress on fluency disorders: Proceedings (pp. 460–465). Nijmegen, Netherlands: Nijmegen University Press.

Gabel, R. M. (2006). Effects of stuttering severity and therapy involvement on attitudes towards people who stutter. *Journal of Fluency Disorders, 31*(3), 216–227. doi: 10.1016/j.jfludis.2006.05.003

Hughes, C. D., Gabel, R. M., Goberman, A. M., & Hughes, S. (2011). Family experiences of people who stutter. *Canadian Journal of Speech-Language Pathology and Audiology, 35*(1), 45–55. Retrieved from http://cjslpa.ca/files/2011_CJSLPA_Vol_35/No_01_1-102/hughes_gabel_goberman_hughes_CJSLPA_2011.pdf

Hughes, C., Gabel, R., & Palasik, S. (2011). Talking about stuttering with a known person who stutters: Impact on perceptions toward stuttering. *Perspectives on Fluency and Fluency Disorders, 21*, 50–58. doi: 10.1044/ffd21.2.50

Hughes, S., Gabel, R., Irani, F., & Schlagheck, A. (2010). University students' perceptions of the life effects of stuttering. *Journal of Communication Disorders, 43*, 45–60. doi: 10.1016/j.jcomdis.2009.09.002

Klassen, T. R. (2001). Perceptions of people who stutter: Re-assessing the negative stereotype. *Perceptual and Motor Skills, 92*(2), 551–559. doi: 10.2466/pms.2001.92.2.551

Klassen, T. R. (2002). Social distance and the negative stereotype of people who stutter. *Journal of Speech-Language Pathology and Audiology, 26*(2), 90–99. Retrieved from http://cjslpa.ca/files/2002_JSLPA_Vol_26/No_02_77-120/Klassen_JSLPA_2002.pdf

Panico, J., Healey, E. C., Brouwer, K., Susca, M. (2005). Listener perceptions of stuttering across two presentation modes: A quantitative and qualitative approach. *Journal of Fluency Disorders, 30*, 65–85. doi: 10.1016/j.jfludis.2005.01.003

Reitzes, P. (2006). Providing mentors, role models, and peer support for children who stutter. *The Journal of Stuttering Therapy, Advocacy & Research, 1*, 16–28.

Schlagheck, A., Gabel, R., & Hughes, S. (2009). A mixed methods study of stereotypes of people who stutter. *Contemporary Issues in Communication Science and Disorders, 36*, 108–117. Retrieved from http://www.asha.org/uploadedFiles/asha/publications/cicsd/2009FMixedMethodsStudy.pdf

Silverman, F. H., & Bongey, T. A. (1997). Nurses' attitudes toward physicians who stutter. Journal of Fluency Disorders, 22, 61–62. doi: 10.1016/S0094-730X(96)00056-3

St. Louis, K. O. (2011). The Public Opinion Survey of Human Attributes - Stuttering (POSHA-S): Summary framework and empirical comparisons. *Journal of Fluency Disorders, 36*(4), 256–261. doi: 10.1016/j.jfludis.2011.02.003

Turnbaugh, K., Guitar, B., & Hoffman, P. (1979). Speech clinicians' attribution of personality traits as a function of stuttering severity. *Journal of Speech and Hearing Research, 22*, 37–45. doi: 10.1044/jshr.2201.37

Walker, R., Mayo, R., & St. Louis, K. O. (2016). Attitudes of college career counselors toward stuttering and people who stutter. *Perspectives of the ASHA Special Interests Groups, 1*(1), 44–53. doi: 10.1044/persp1.SIG4.44

Woods, C. L. & Williams, D. (1976). Traits attributed to stuttering and normally fluent males. *Journal of Speech and Hearing Research, 36*, 267–278. doi: 10.1044/jshr.1902.267

Yairi, E., & Carrico, D. M. (1992). Early childhood stuttering: Pediatricians' attitudes and practices. *American Journal of Speech-Language Pathology, 1*, 54–62. doi: 10.1044/1058-0360.0103.54

Yaruss, J. S. & Quesal, R. W. (2004). Stuttering and the International Classification of Functioning, Disability, and Health (ICF): An update. *Journal of Communication Disorders, 37*(1), 35–52. doi: 10.1016/S0021-9924(03)00052-2

Acknowledgements

The authors would like to thank the participants for taking the time to participate in this study.

Authors' Note

The authors do not have any conflicts of interest to disclose related to this study. Please address any correspondence related to this article to Charles D. Hughes at the Department of Communication Sciences and Disorders, 200 Health and Human Services Building, Bowling Green State University, Bowling Green, Ohio, 43403. Email: **chughes@bgsu.edu**.

Performance of Young, Middle-Aged, and Older Adults on Tests of Executive Function

La performance des jeunes adultes, des adultes d'âge moyen et des aînés à des tests évaluant les fonctions exécutives

KEY WORDS

EXECUTIVE FUNCTION

COGNITION

AGING

Angela N. Burda
Emily Andersen
Marissa Berryman
Maddisen Heun
Claire King
Tina Kise

Abstract

Information on differently aged adults' performance on tests of executive function administered by speech-language pathologists is lacking. This potentially limits clinicians' abilities to accurately evaluate and treat persons with cognitive impairments. The objective of this study was to determine potential differences among young, middle-aged, and older adults on 2 tests of executive function: the Behavioural Assessment of Dysexecutive Syndrome and the Functional Assessment of Verbal Reasoning and Executive Strategies. In total, 105 healthy adult participants completed both tests in this pilot study. Participants were equally divided into the following 3 age groups: Young, Middle-aged, and Older, with ages ranging from 20–88 years old. Older adults demonstrated statistically significantly lower scores compared to young and middle-aged adults on both tests. No significant performance differences were found between young and middle-aged adults. Further research is necessary to determine a definitive pattern of performance on these tests in adults across the lifespan.

Angela N. Burda
Emily Andersen
Marissa Berryman
Maddisen Heun
Claire King
Tina Kise
University of Northern Iowa, Cedar Falls, Iowa, USA

Abrégé

L'information concernant la performance des adultes de différentes tranches d'âge à des tests évaluant les fonctions exécutives et administrés par les orthophonistes est manquante. Cette situation peut limiter la capacité des cliniciens à évaluer avec précision et à intervenir auprès de personnes ayant un trouble cognitif. L'objectif de cette étude était de déterminer les différences potentielles entre les performances des jeunes adultes, des adultes d'âge moyen et des aînés à deux tests évaluant les fonctions exécutives : le *Behavioural Assessment of Dysexecutive Syndrome* et le *Functional Assessment of Verbal Reasoning and Executive Strategies*. Au total, 105 adultes en santé ont complété les deux tests de cette étude pilote. Les participants ont été divisés en trois groupes égaux en fonction de leur âge : jeunes adultes, adultes d'âge moyen et aînés. L'âge des participants variait entre 20 et 88 ans. Les aînés ont obtenu des résultats significativement plus faibles aux deux tests comparativement aux jeunes adultes et aux adultes d'âge moyen. Aucune différence significative n'a été trouvée entre les performances des jeunes adultes et celle des adultes d'âge moyen. Des recherches supplémentaires sont nécessaires afin de déterminer les profils de performance des adultes à ces tests, et ce, aux différents âges de la vie.

Executive function is a term that encompasses numerous abilities involving higher level cognitive processes, including: initiating, forming goals, applying knowledge and judgment in problem-solving situations, sequencing, carrying out plans to completion, inhibiting inappropriate behaviours, and organizing pertinent information (Crawford & Channon, 2002; Pickens, Ostwald, Murphy-Pace, & Bergstrom, 2010). In essence, intact executive functioning facilitates dynamic adaptations to novel and varied situations. Impaired executive function can adversely impact the completion of daily activities, social communication, and social cognition. Persons with executive dysfunction can lack structure and coherence in discourse and leave out pertinent information during conversation (Douglas, 2010). Individuals can also have difficulty interpreting the behaviour of others and have reduced theory of mind (Sohlberg & Turkstra, 2011; Van Overwalle, Baetens, Mariën, & Vandekrerckhove, 2014). Such challenges can make interactions and conversations with others difficult (Douglas, 2010; Sohlberg & Turkstra, 2011).

Executive function performance in differently aged adults

Some researchers have reported executive function performance decreases with age and declines earlier than previously believed (Allain et al., 2005; Garden, Phillips, & MacPherson, 2001). Others have reported little executive function decline until old age, and that cognitive declines in those under 60 years of age are not typically clinically important (Singh-Manoux et al., 2012). However, individuals with Alzheimer's disease have shown slight cognitive changes 10–20 years prior to diagnosis (Rajan, Wilson, Weuve, Barnes, & Evans, 2015; Tondelli et al., 2012). If older adults develop executive dysfunction, their functional status may be affected (Pickens et al., 2010) as deficits in pragmatics, discourse, memory, attention, and strategic thinking typically occur (Geffner, 2007). The ability to make decisions autonomously may be called into question when individuals display characteristics of executive dysfunction (Pickens et al., 2010), possibly impacting the capacity to live independently or the ability to provide informed consent for a medical procedure or care.

Young and middle-aged adults have performed well on tasks measuring executive function, including those that mimic the real world and require open-ended planning (Allain et al., 2005; Garden et al., 2001). More recently, Burda et al. (2014) examined performance differences between healthy younger and middle-aged adults on two tests of executive function: the Behavioural Assessment of Dysexecutive Syndrome (BADS; Wilson, Alderman, Burgess, Emslie, & Evans, 1996) and the Functional Assessment of Verbal Reasoning and Executive Strategies (FAVRES; MacDonald, 2005). Middle-aged adults had higher scores than young adults on the Rule Shift task of the BADS. No other differences occurred on the remaining subtests or any subtests on the FAVRES. Older adults were not included in that study.

Tests to evaluate executive function performance

Tests of executive function require individuals to perform tasks that evaluate various skills (Purdy, 2015). Faria, Alves, and Charchat-Fichman (2015) recently reported some of the most frequently used neuropsychological tests to evaluate executive functions in older adults were the Trail Making Test (TMT) Form B; the Verbal Fluency Test (VFT) – F, A, S and the Animals category; the Clock Drawing Test (CDT); and the Stroop Test (Lezak, Howieson, Bigler, & Tranel, 2012). These tests have similarities to the BADS and FAVRES, the two tests used in this study. The TMT Form B and portions of the BADS require individuals to use working memory and repeatedly switch attention between different sequences. The VFT tasks require individuals to search their memory for specific information; semantic memory is also assessed. The FAVRES evaluates semantic memory during generation tasks that are comparable to verbal fluency tasks. The CDT is a visuospatial pen and paper task that requires planning within an allotted space, similar to the Key Search on the BADS. The Stroop Test and Test 1 of the BADS both evaluate inhibitory control. Although the FAVRES may not have as many tasks that directly match those of the tests reported in Faria et al. (2015), its subtests better reflect everyday activities (e.g., planning one's work day or writing a letter of complaint; MacDonald & Johnson, 2005). Similar to Allain et al. (2005), the majority of studies that utilized the tests of executive function discussed in Faria et al. (2015) found that older adults tended to have lower scores versus younger adults. Many studies also included more than one test since different tests evaluate different executive function abilities (Faria et al., 2015).

Of the tests used in this study, the BADS includes six tests (i.e., subtests) that determine the severity of dysexecutive impairments by evaluating high-level tasks such as "planning, organising, initiating, monitoring and adapting behaviour" (Chamberlain, 2003, p. 33). Individuals provide verbal and written responses and complete a hands-on activity. One test asks temporal judgment questions (e.g., How long do most dogs live

for?). Another test provides persons with a zoo map and open-ended instructions to visit several exhibits following a set of rules. After completing this task, patients are given the same zoo map with more specific instructions on the sequence of exhibits to visit. The normative sample was composed of 216 healthy adults grouped into age brackets (i.e., 16–31, 32–47, 48–63, 64 and older) and 78 persons aged 19–78 years with various neurological disorders. Although participants in the norming group under the age of 64 performed significantly better than those aged 64 or older, no comparisons between young and middle-aged adults were included (Wilson et al., 1996).

The FAVRES assesses four high-level cognitive-communication skills that can occur in daily life: planning an event, scheduling a workday, making a decision, and building a case (MacDonald & Johnson, 2005, p. 896). Planning, organizing, sequencing, controlling inhibitions, and "prioritizing tasks with time constraints" are assessed (MacDonald & Johnson, 2005, p. 897). Tasks (i.e., subtests) generally contain restrictions (e.g., meetings must occur at specific times when scheduling a workday). Generation and prediction tasks are also completed. For example, after planning a children's event in Task 1, patients generate activities one could do with an adult and then predict two good and two bad things that could happen at the chosen event. As opposed to laboratory measures, ecologically valid tasks can give a better idea of daily functioning (Moriyama et al., 2002; Sussman, Rychtarik, Mueser, Glynn, & Pruesu, 1986), possibly helping to predict individuals' behaviours in daily life (Silver, 2000). The FAVRES was normed on 101 healthy adults ages 17–89 years and 52 adults with an acquired brain injury; no information was included on age-related performance (MacDonald, 2005).

Objective of the study

Speech-language pathologists (S-LPs) work with several populations who exhibit executive dysfunction (e.g., persons with brain injuries, multiple sclerosis, or dementia; Geffner, 2007; Royall, Palmer, Chiodo, & Polk, 2004). Clinical assessments are generally based on traditional tasks rather than functional tasks representative of real life, allowing for gross misestimates of performance (Crawford & Channon, 2002). The relatively sparse normative data on tests of executive function that S-LPs may use complicates the matter. The literature lacks specific information on potential performance differences between differently aged adults on the BADS and FAVRES (MacDonald, 2005; Wilson et al., 1996). While Burda et al. (2014) reported little performance difference on the BADS and FAVRES between young and middle-aged adults, data on the performance of older adults on these specific tests is lacking. Yet, medically based S-LPs need to know how well healthy adults across the lifespan perform on these tests in order to determine if their patients' performance is indicative of cognitive-communicative deficits or if their performance is age-appropriate. Such information could further aid treatment and prognosis by providing a clearer picture of how much cognitive change can be attributed to normal aging. The current study extended the study by Burda et al. (2014) by including older adults. The objective of this pilot study addressed the following research question: Are there statistically significant differences between young, middle-aged, and older adults on the BADS and the FAVRES?

Methods

Participants

Following approval of the protocol by the University of Northern Iowa's Institutional Review Board (Protocol #: 09-0270), participants were recruited for this cross-sectional quasi-experimental study from small, mid-sized, and large urban and rural communities in the Midwest by posting flyers in public areas (e.g., libraries). A power analysis for an effect size of .08 with an alpha of .05 indicated that a total sample size of 105 was needed. Participants were equally divided into the following age groups: Young (aged 20–39 years), Middle-Aged (aged 40–59 years), and Older (aged 60 and older). Participant inclusion criteria included: no history of any neurological damage or events, possessing at least a high school level of education, native English-speaking, and passing a pure tone hearing screening with tones presented at 20 dB HL at the frequencies of 500, 1000, 2000, and 4000 Hz for the young and middle-aged adults. Older participants were included if they had no greater than a mild hearing loss in their better ear, defined as no greater than 40 dB hearing loss at any of the previously documented frequencies (Burda, Casey, Foster, Pilkington, & Reppe, 2006). Participants were required to score a minimum of 28 or higher on the *Mini-Mental State Examination* (MMSE; Folstein, Folstein, & Fanjiang, 2001). Basic ethical considerations adopted by the University of Northern Iowa were taken to ensure the protection of participants in this study.

Stimuli and procedures

Tests were administered according to the test manual protocols. In order to control for possible testing order effects, every other participant (n = 53) was administered the BADS (Wilson et al., 1996) first; the remaining 52 participants were administered the FAVRES (MacDonald,

2005) first. This process was followed for each age group. Testing was completed in a single session, typically lasting 120 minutes. Breaks were provided as needed.

Data analysis

Participant responses were scored according to the procedures found in the test manuals. Raw scores on the BADS were converted to profile scores. Profile scores ranged from 0–4. Some subtests (e.g., the Modified Six Elements Test) had timed components, which factored into calculating the profile score. Performance of normal controls indicated planning time and time to complete a task were essential elements of executive function. Summing the profile scores for each of the six tests led to an overall profile score. If patients completed the entire test, they earned a Total Profile Score ranging from 0–24.

Participants earned the following raw scores for each FAVRES Task (i.e., subtest): Time, Accuracy, and Rationale. They also earned raw scores on Reasoning Subskills. Individuals earned the highest points possible for the most appropriate response. If participants provided a reasonable related response, they earned some, but not all, of the points. Raw scores were then converted to standard scores in the same areas (e.g., Time, Accuracy). Raw scores were also used to calculate the Total Score for the test. The mean standard score was 100 with a standard deviation of 15.

To address the study's objective, a series of two-factor Analyses of Variance (ANOVAs) with Bonferroni correction for multiple comparisons were used. Age was the independent variable and Score was the dependent variable. Since higher levels of education have led to higher scores on tasks measuring executive function (Ardila, Ostrosky-Solis, Rosselli, & Gomez, 2000), an additional series of two-factor ANOVAs were conducted to determine potential significant differences between participant groups' education levels and MMSE scores. Age was the independent variable; Education Level and Score were the dependent variables, respectively. Statistical analyses were performed using SPSS 22.0.

Reliability

Subtest raw scores were used to calculate inter- and intra-rater reliability on approximately 20% of a randomly chosen sample (i.e., 15 participants). For inter-rater reliability, the investigators' scores were correlated with scores of a trained speech-language pathology graduate student. For intra-rater reliability, the investigators scored the selected protocols twice. The second scoring took place two weeks after the initial scoring. Pearson *r* correlations were calculated. Inter-rater reliability for the FAVRES was $r = .92$; intra-rater reliability was $r = .94$. Inter-rater reliability for the BADS was $r = .90$; intra-rater reliability was $r = .94$.

Results

Participants

Participants were 105 adults (49 men, 56 women) with 35 participants in each age group. Participants had high mean MMSE scores. The majority had completed or were completing some type of post-high school education (See Table 1).

Table 1. Demographics

	Young Adults (*n* = 35)		Middle-Aged Adults (*n* = 35)		Older Adults (*n* = 35)	
Demographic Information	*M*	*SD*	*M*	*SD*	*M*	*SD*
Age in Years	23.71	5.25	50.31	5.27	69.83	8.21
MMSE Score*	29.00	1.41	29.46	.78	28.34	1.59
Years of College	6.66	2.84	3.06	2.22	5.03	3.16

Note. MMSE = Mini-Mental State Examination. *The highest possible score on the MMSE is 30.

Descriptive statistics

Overall mean scores on the BADS and the FAVRES were obtained (see Tables 2 and 3). In both tests, older adults generally had the lowest mean scores compared to the other two groups. There were some exceptions. On the BADS, middle-aged adults had the highest mean scores for the Action Program while young adults had the lowest mean scores. On the FAVRES, older adults had the lowest mean standard scores for Accuracy and Rationale measures. For measures of Time, young adults had the lowest mean standard scores for Task 1 (Planning an Event), while middle-aged adults had the lowest mean scores for Task 2 (Scheduling).

Inferential statistics

Several statistically significant differences occurred in both tests. On the BADS, the results showed an effect of group for the Rule Shift Card test, F (2, 104) = 5.46, $p \leq .006$; the Zoo Map test, F (2, 104) = 4.65, $p \leq .01$; the BADS Total Profile Score, F (2, 104) = 6.34, $p \leq .003$; and the BADS Standard Score, F (2, 104) = 6.22, $p \leq .003$.

On the FAVRES, an effect for group occurred for Accuracy scores on the following subtests: Task 1 (Planning an Event), F (2, 104) = 4.41, $p \leq .014$; Task 2 (Scheduling), F (2, 104) = 8.91, $p \leq .0001$; Task 4 (Building a Case), F (2, 104) = 4.69, $p \leq .01$; and Accuracy Total, F (2, 104) = 7.64, $p \leq .001$. Significant differences also occurred for Rationale

Table 2. Mean and Standard Deviations of BADS Profile Scores for Young, Middle-Aged, and Older Adults

		Young Adults (n = 35)		Middle-Aged Adults (n = 35)		Older Adults (n = 35)	
BADS Subtests	Total Score Possible	*M*	*SD*	*M*	*SD*	*M*	*SD*
Rule-Shift Cards	4	3.40	.65	3.80	.47	3.23	1.00
Action Program	4	3.46	1.22	3.89	.53	3.60	.81
Key Search	4	3.26	.95	2.86	1.17	2.77	1.00
Temporal Judgment	4	1.20	.63	1.31	.68	1.17	.57
Zoo Map	4	2.89	.99	2.51	1.09	2.09	1.20
Modified Six Elements	4	3.66	.76	3.74	.89	3.40	.81
Total Points Score	24	17.89	1.95	17.97	2.26	16.26	2.56
Standard Score	100	98.97	9.34	99.37	10.69	91.29	12.19

Note. BADS = Behavioural Assessment of Dysexecutive Syndrome.

Table 3. Means and Standard Deviations of FAVRES Standard Scores for Young, Middle-Aged, and Older Adults

		Young Adults (n = 35)		Middle-Aged Adults (n = 35)		Older Adults (n = 35)	
FAVRES Tasks Accuracy SS	Total SS Possible	*M*	*SD*	*M*	*SD*	*M*	*SD*
Task 1	108	100.94	16.02	96.89	18.43	87.91	20.58
Task 2	106	90.46	21.26	95.11	17.09	73.31	32.79
Task 3	107	93.57	22.42	71.89	30.39	90.23	29.64
Task 4	106	82.00	30.77	85.80	19.48	57.31	39.73
Total Test	111	88.40	21.22	97.23	21.34	65.91	35.36
Rationale SS							
Task 1	106	101.77	14.94	95.37	19.50	87.31	31.84
Task 2	109	99.31	14.94	89.74	28.09	88.74	18.50
Task 3	103	92.60	22.03	82.20	28.26	72.31	39.94
Task 4	107	80.11	31.89	88.37	17.98	67.17	33.55
Total Test	111	89.20	20.85	109.60	9.54	68.09	27.17
Time SS							
Task 1	132	105.91	16.09	93.83	25.28	105.97	12.20
Task 2	144	101.23	18.99	100.11	13.78	94.06	25.00
Task 3	130	104.51	9.82	105.77	10.60	94.49	21.43
Task 4	135	109.00	11.77	104.06	11.71	105.54	11.78
Total Test	126	107.89	14.82	86.03	31.96	101.03	17.91
Reasoning SS							
	142	91.09	13.42	86.94	32.10	82.71	13.96

Note. FAVRES = Functional Assessment of Verbal Reasoning and Executive Strategies; SS = Standard Score.

scores for the following: Task 1, F (2, 104) = 3.39, $p \leq .04$; Task 2, F (2, 104) = 3.17, $p \leq .05$; Task 3 (Making a Decision), F (2, 104) = 4.41, $p \leq .02$; and Rationale Total, F (2, 104) = 10.04, $p \leq .0001$. Finally, statistically significant differences occurred for Task 3 Time scores, F (2, 104) = 3.56, $p \leq .03$.

Post hoc testing. Tukey's post hoc testing on the BADS indicated middle-aged adults had statistically significantly higher scores on the Rule Shift Card test compared to older adults (Tukey's Value = 0.57, $p \leq .005$); young adults scored significantly higher on the Zoo Map task than did older adults (Tukey's Value = 0.80, $p \leq .008$). For the BADS Total Profile score, both young and middle-aged adults had significantly higher scores on the BADS Total Profile Score versus older adults (Tukey's Value = 1.62, $p \leq .009$ and Tukey's Value = 1.71, $p \leq .006$, respectively). Young and middle-aged adults also had statistically higher scores on the BADS Standard Score compared to the older adults (Tukey's Value = 7.68, $p \leq .01$ and Tukey's Value = 8.09, $p \leq .006$, respectively).

Results from Tukey's post hoc testing on the FAVRES revealed that for Accuracy scores, young adults had statistically significantly higher scores on Task 1 than older adults (Tukey's Value = 13.02, $p \leq .011$). Young and middle-aged adults scored higher on Task 2 than older adults (Tukey's Value = 18.71, $p \leq .006$ and Tukey's Value = 23.57, $p \leq .000$, respectively). Young adults also scored significantly higher on Task 4 compared to older adults (Tukey's Value = 24.68, $p \leq .008$). Young and middle-aged adults had significantly higher Accuracy Total scores than older adults (Tukey's Value = 22.48, $p \leq .002$ and Tukey's Value = 19.88, $p \leq .006$, respectively). For Rationale scores, Tukey's post hoc testing indicated that young adults had statistically significantly higher scores than older adults on Tasks 1, 2, and 3 (Tukey's Value = 14.46, $p \leq .033$; Tukey's Value = 10.57, $p \leq .038$; and Tukey's Value = 20.29, $p \leq .019$, respectively). Young and middle-aged adults had significantly higher Rationale Total scores compared to older adults (Tukey's Value = 21.11, $p \leq .000$ and Tukey's Value = 20.28, $p \leq .001$, respectively). Finally, young adults had higher Task 3 Time scores than older adults (Tukey's Value = 10.02, $p \leq .024$).

Education levels and MMSE scores. No statistically significant differences occurred among age groups for education levels or MMSE scores, F(2, 104) = 1.33, $p \geq .27$ and F(2, 104) = 0.17, $p \geq .92$ respectively.

Discussion

This pilot study is one of few that have investigated young, middle-aged, and older healthy adults' performance on the BADS and the FAVRES. The current study adds information that was previously unavailable in both tests. Overall, age appeared to affect performance on both the BADS and the FAVRES. As hypothesized, older adults had statistically significantly lower scores compared to young and middle-aged adults on several subtests. Not surprisingly, no significant differences occurred between young and middle-aged adults. Burda et al. (2014) found young and middle-aged adults had no significant performance differences on the FAVRES and all but one subtest on the BADS. Middle-aged adults had significantly higher scores on the Rule Card Shift Test versus young adults. Garden et al. (2001) also found no evidence of middle-aged adults having difficulty with changing tasks or following rules on a task similar to the six elements subtest of the BADS.

In the current study, older adults had the majority of the lowest mean scores on both the BADS and the FAVRES, likely one of the most ecologically valid executive function tests available to S-LPs. An interesting trend is that there were no significant performance differences on Task 3 (Making a Decision) among age groups, although younger adults took more time completing the task compared to the other groups. It is uncertain why this particular subtest did not garner similar outcomes (i.e., older adults having significantly lower scores vs. the other groups). Results of this study concur with the assertions by Allain et al. (2005) and Garden et al. (2001) that executive function performance decreases with age. The results also mirror findings by Allain et al. (2005) in that older adults had poorer performance on the BADS Zoo Map test compared to young adults. Such results may not be surprising because, compared to other cognitive tests, tests of executive function can be more sensitive to the effects of aging due to their complexity (Morris, Worsley, & Matthews, 2000; Murray, 2012). However, all older participants self-reported no history of neurological events (e.g., transient ischemic attack), and all were living on their own at the time of testing. Consequently, findings from this study must be interpreted cautiously. While this study adds to the literature, a broad statement denoting that lower scores on the BADS and FAVRES are typical of healthy older adults cannot be made until more research indicates this is indeed the case. In addition, care should be taken when interpreting scores from the BADS and FAVRES in clinical settings. Older adult patients may have performed more poorly on these tests pre-morbidly than younger or middle-aged patients. Thus, further inquiry may be necessary to ascertain if older patients have executive dysfunction and if so, to what extent. Interviews with patients and/or loved ones could aid in determining pre-morbid level of functioning and evidence of potential cognitive declines. Careful comparison of test

scores with these responses and observations will ideally allow S-LPs to develop treatment plans that best meet their patients' needs.

Limitations and future research

Limitations exist with this investigation. While Singh-Manoux et al. (2012) reported that cognitive declines in individuals under 60 years of age are usually not clinically important, such assertions require longitudinal study. Thus, no predictions can be made based on results of the current study. However, participants could become familiar with tests in a longitudinal study, potentially biasing results (Singh-Manoux et al., 2012). Lack of randomization led to participants who were generally highly educated and skewed to the lower end of age ranges, particularly the older adults (M_{age} = 69.83). Few participants represented minority populations, limiting generalization of the current study's results (Scheffner Hammer, 2011). Adults from culturally and linguistically diverse backgrounds with neurologically based acquired communication disorders may perform differently on these tests (Ellis, 2009; Scheffner Hammer, 2011).

Future research options include testing a randomized participant pool that represents a variety of diverse populations and education levels. Forthcoming investigations could include aspects such as participants' physical activity, diet, and mental and social engagement. Previous studies have noted that low physical activity, high saturated fat intake, high dietary cholesterol, and a lack of mental and social engagement negatively affected cognitive abilities, including executive function, in adults across the lifespan (Fratiglioni, Paillard-Borg, & Winblad, 2004; Morris & Tangney, 2014; Singh-Manoux, Hillsdon, Brunner, & Marmot, 2005). Further research is needed to determine how differently aged individuals with acquired neurogenic communication disorders perform on the BADS and FAVRES compared to healthy age-matched controls. While more data must be obtained, S-LPs should be aware that healthy older adults could evidence lower scores on the BADS and FAVRES compared to younger and middle-aged adults.

References

Allain, P., Nicoleau, S., Pinon, K., Etcharry-Bouyx, F., Barré, J., Berrut, G., & Gall, D. L. (2005). Executive functioning in normal aging: A study of action planning using the Zoo Map Test. *Brain and Cognition, 57*, 4–7. doi: 10.1016/j.bandc.2004.08.011

Ardila, A., Ostrosky-Solis, F., Rosselli, M., & Gomez, C. (2000). Age related cognitive decline during normal aging: The complex effect of education. *Archives of Clinical Neuropsychology, 15*, 495–514. doi: 10.1016/S0887-6177(99)000040-2

Burda, A., Baker, K., Burns, A., Butler, C., Clark, L., Cook, W., ... Young, M. (2014). Performance of young and middle-aged adults on cognitive tests. *Journal of Medical Speech-Language Pathology, 21*, 393–404.

Burda, A. N, Casey, A. M., Foster, T. R., Pilkington, A. K., & Reppe, E. A. (2006). Effects of accent and age on transcription of medically related utterances: A pilot study. *Communication Disorders Quarterly, 27*, 110–116. doi: 10.1177/2F15257401060270020101

Chamberlain, E. (2003). Behavioral Assessment of the Dysexecutive Syndrome (BADS). [Test Review]. *Journal of Occupational Psychology, Employment and Disability, 5*, 33–37.

Crawford, S., & Channon, S. (2002). Dissociation between performance on abstract tests of executive function and problem solving in real-life-type situations in normal aging. *Aging & Mental Health, 6*, 12–21. doi: 10.1080/13607860120101130

Douglas, J. M. (2010). Relation of executive functioning to pragmatic outcome following severe traumatic brain injury. *Journal of Speech Language and Hearing Research, 17*, 365–382. doi: 10.1044/1029-4388

Ellis, C. (2009). Does race/ethnicity really matter in adult neurogenics? *American Journal of Speech-Language Pathology, 18,* 310–314. doi: 10.1044/1058-0360(2009/08-0039)

Faria, C. D. A., Alves, H. V. D., & Charchat-Fichman, H. (2015). The most frequently used tests for assessing executive functions in aging. *Dementia & Neuropsychologia, 9*(2), 149–155. doi: 10.1590/1980-57642015DN92000009

Folstein, M. F., Folstein, S. E., & Fanjiang, G. (2001). *Mini-Mental State Examination: Clinical guide*. Lutz, FL: Psychological Assessment Resources.

Fratiglioni, L., Paillard-Borg, S., & Winblad, B. (2004). An active and socially integrated lifestyle in late life might protect against dementia. *The Lancet, 3*, 343–353. doi: 10.1016/S1474-4422(04)00767-7

Garden, S. E., Phillips, L. H., & MacPherson, S. E. (2001). Midlife aging, open-ended planning, and laboratory measures of executive function. *Neuropsychology, 15*, 472–482. doi: 10.1037/0894-4105.15.4.472

Geffner, D. (2007, November). *Managing executive function disorders*. Presentation at the Annual Convention of the American Speech-Language-Hearing Association, Boston, MA.

Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological assessment* (5th ed.). Oxford, UK: Oxford University Press.

MacDonald, S. (2005). *Function Assessment of Verbal Reasoning and Executive Strategies*. Guelph, Canada: CCD Publishing.

MacDonald, S., & Johnson, C. J. (2005). Assessment of subtle cognitive-communication deficits following acquired brain injury: A normative study of the Functional Assessment of Verbal Reasoning and Executive Strategies (FAVRES). *Brain Injury, 19*, 895–902.

Moriyama, Y., Mimura, M., Kato, M., Yoshino, A., Hara, T., Kashima, H., ... Watanabe, A. (2002). Executive dysfunction and clinical outcome in chronic alcoholics. *Alcoholism: Clinical and Experimental Research, 26*, 1239–1244. doi: 10.1097/01.ALC.0000026103.08053.86

Morris, M. C., & Tangney, C. C. (2014). Dietary fat composition and dementia risk. *Neurobiology of Aging, 35* (Suppl. 2), S59– S64. doi: 10.1016/j.neurobiolaging.2014.03.038

Morris, R. G., Worsley, C., & Matthews, D. (2000). Neuropsychological assessment in older people: Old principles and new directions. *Advances in Psychiatric Treatment, 6*, 362–372. doi: 10.1192/apt.6.5.362

Murray, L. (2012). Assessing cognitive function in older patients: The why, who, what, and how. *Perspectives on Gerontology, 17*, 17–26. doi: 10.1044/gero17.1.17

Pickens, S., Ostwald, S. K., Murphy-Pace, K., & Bergstrom, N. (2010). Evidence synthesis: Systematic review of current executive function measures in adults with and without cognitive impairments. *International Journal of Evidence-Based Healthcare, 8*, 110–125. doi: 10.1111/j.1744-1609.2010.00170.x

Purdy, M. H. (2015). Executive functions: Theory, assessment, and treatment. In M. L. Kimbarow (Ed.), *Cognitive communication disorders* (2nd ed.; pp. 83-128). San Diego, CA: Plural Publishing, Inc.

Rajan, K. B., Wilson, R. S., Weuve, J., Barnes, L. L., & Evans, D. A. (2015). Cognitive impairment 18 years before clinical diagnosis of Alzheimer disease dementia. *Neurology, 85*, 898–904. doi: 10.1001/jamaneurol.2014.3375

Royall, D. R., Palmer, R., Chiodo, L. K., & Polk, M. J. (2004). Declining executive control in normal aging predicts change in functional status: The freedom house study. *Journal of the American Geriatrics Society, 52*, 346-352. doi: 10.1111/j.1532-5415.2004.52104.x

Scheffner Hammer, C. (2011). Broadening our knowledge about diverse populations. *American Journal of Speech-Language Pathology, 20*, 71–72. doi: 10.1044/1058-0360(2011/ed-02)

Silver, C. H. (2000). Ecological validity of neuropsychological assessment in childhood traumatic brain injury. *Journal of Head Trauma Rehabilitation, 15*, 973–988. doi: 10.1097/00001199-200008000-0002

Singh-Manoux, A., Hillsdon, M., Brunner, E., & Marmot, M. (2005). Effects of physical activity on cognitive functioning in middle age: Evidence from the Whitehall II prospective cohort study. *American Journal of Public Health, 95*, 2252-2258. doi: 10.2105/AJPH.2004.055574

Singh-Manoux, A., Kivimaki, M., Glymour, M. M., Elbaz, A., Berr, C., Ebmeier, K. P., … Dugravot, A. (2012). Timing of onset of cognitive decline: Results from Whitehall II prospective cohort study. *British Medical Journal, 344*, 1–8. doi: 10.1136/bmj.d7622

Sohlberg, M. M. & Turkstra, L. (2011). *Optimizing cognitive rehabilitation: Effective instructional methods*. New York, NY: The Guilford Press.

Sussman, S., Rychtarik, R. G., Mueser, K., Glynn, S., & Pruesu, D. M. (1986). Ecological relevance of memory tests and the prediction of relapse in alcoholics. *Journal of Studies on Alcohol and Drugs, 47*, 305–310. doi: 10.15288/jsa.1986.47.305

Tondelli, M., Wilcock, G. K., Nichelli, P., De Jager, C. A., Jenkinson, M., & Zamboni, G. (2012). Structural MRI changes detectable up to ten years before clinical Alzheimer's disease. *Neurobiology of Aging, 33*, e825–e836. doi: 10.1016/j.neurobiolaging.2011.05.018

Van Overwalle, F., Baetens, K., Mariën, P., & Vandekrerckhove, M. (2014). Social cognition and the cerebellum: A meta-analysis of over 350 fMRI studies. *Neuroimage, 86*, 554–572. doi:10.1016/j.neuroimage.2013.09.033

Wilson, B. A., Alderman, N., Burgess, P. W., Emslie, H., & Evans, J. J. (1996). *Behavioural Assessment of the Dysexecutive Syndrome*. London, England: Thames Valley Test Company.

Acknowledgements

Funding was provided by the University of Northern Iowa's College of Humanities, Arts, and Sciences Small Grant.

Authors' Note

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

Correspondence concerning this article should be addressed to Angela N. Burda, Department of Communication Sciences and Disorders, University of Northern Iowa, 230 Communication Arts Center, Cedar Falls (IA), 50614-0356, U.S.A. Email: angela.burda@uni.edu.

Canadian-French Validation of Two Questionnaires That Measure the Stigma Associated With Hearing Impairment: Initial Development

Développement initial et validation franco-canadienne de deux questionnaires mesurant les stigmates associés à une perte auditive

KEY WORDS

STIGMA QUESTIONNAIRE
HEARING LOSS
HEARING AIDS
INSTRUMENT TRANSLATION
TRANSCULTURAL VALIDATION
PSYCHOMETRIC TESTING

Claude Vincent
Jean-Pierre Gagné
Tony Leroux
Audrey Clothier
Marianne Larivière
Frédéric S. Dumont
Martine Gendron

Claude Vincent
Université Laval,
Québec, QC
CANADA
Centre interdisciplinaire de recherche en réadaptation et intégration sociale,
Québec, QC
CANADA

Jean-Pierre Gagné
Université de Montréal,
Montréal, QC
CANADA

Tony Leroux
Université de Montréal,
Montréal, QC
CANADA
Centre de recherche interdisciplinaire en réadaptation du Montréal métropolitain,
Montréal, QC
CANADA

Audrey Clothier
Centre interdisciplinaire de recherche en réadaptation et intégration sociale,
Québec, QC
CANADA

Marianne Larivière
Université de Montréal,
Montréal, QC
CANADA

Frédéric S. Dumont
Centre interdisciplinaire de recherche en réadaptation et intégration sociale,
Québec, QC
CANADA

Martine Gendron
Centre de recherche interdisciplinaire en réadaptation du Montréal métropolitain,
Montréal, QC
CANADA

Abstract

Questionnaires evaluating stigma and its consequences are available in English for several stigmatizing traits. In many Western societies, including French-speaking countries, hearing loss is a stigmatizing trait. Hence, there is a need for French-language standardized questionnaires for measuring stigma associated with hearing loss. The goal of this study was to adapt, translate, and validate 2 questionnaires that assess different aspects of stigmatization and its consequences among adults with hearing impairment. The Stigma Consciousness Questionnaire (SCQ) for Women and the Stigma Scale for Mental Illness were specifically adapted for older adults with hearing impairment. The strategy consisted of the translation and back-translation of the questionnaires by 2 translators, revision by a committee of experts, and administration to 5 bilingual older participants. These 2 novel questionnaires were then administered to 32 Canadian-French participants, 65 years of age or older. For the Canadian-French adaptations of the SCQ for Hearing Loss (SCQ-CF) and the Hearing Loss Stigma Questionnaire (HLS-CF), the results yielded good internal consistency (α = .79 and .84, respectively) and slightly lower repeatability, with about 10% (1/10 and 3/28) of the items having no significant test-retest correlations. Factor analysis performed on the SCQ-CF data indicated 3 factors rather than the single factor reported for the original questionnaire. This study resulted in 2 English and French questionnaires for assessing stigma associated with hearing loss that will be used for further validations.

Abrégé

Plusieurs questionnaires évaluant les stigmates, ainsi que les conséquences y étant associées, sont disponibles en anglais pour divers traits stigmatisants. Dans de nombreux pays occidentaux, dont les pays francophones, la perte auditive est un trait stigmatisant. Des questionnaires normalisés en français sont donc nécessaires pour mesurer les stigmates associés à la perte auditive. L'objectif de cette étude était d'adapter, de traduire et de valider deux questionnaires évaluant différents aspects de la stigmatisation, ainsi que les conséquences y étant associées, auprès d'adultes ayant une perte auditive. Deux questionnaires, soit le *Stigma Consciousness Questionnaire (SCQ) for Women* et le *Stigma Scale for Mental Illness*, ont été adaptés spécifiquement pour les aînés ayant une perte auditive. Ces questionnaires ont été traduits en français, puis retraduits en anglais (processus de traduction inversée), par deux traducteurs. Ils ont ensuite été révisés par un comité d'experts et administrés à cinq participants ainés bilingues. Enfin, les deux nouveaux questionnaires ont été administrés à 32 participants franco-canadiens âgés de 65 ans et plus. Les résultats montrent que les adaptations franco-canadiennes des questionnaires *SCQ for Hearing Loss (SCQ-CF)* et *Hearing Loss Stigma Questionnaire (HLS-CF)* ont une bonne cohérence interne (α = 0,79 et 0,84, respectivement) et une stabilité légèrement inférieure à celle des versions originales : environ 10 % (1/10 et 3/28) des items n'ont pas de corrélation significative lors du test-retest. L'analyse factorielle effectuée sur les données du *SCQ-CF* a identifié trois facteurs, alors que la version originale du questionnaire en avait identifié un seul. Cette étude a permis d'obtenir deux questionnaires évaluant les stigmates associés à la perte auditive (disponibles en anglais et en français) et qui feront l'objet de validations supplémentaires.

People with hearing impairment may be perceived as less capable, cognitively diminished, and poor social communication partners (Heine & Browning, 2002; Jennings, Southall, & Gagné, 2013; Kochkin, 2007; Parrette & Scherer, 2004; Southall, Gagné, & Jennings, 2010). In terms of self-perception, hearing difficulties can lead hearing-impaired individuals to consider themselves old, weak, and less capable, leading them to shun rehabilitation services (Gagné, Southall, & Jennings, 2009). This may cause them to believe that others judge them negatively. The social and self-stigma associated with hearing loss constitutes one of the most important barriers to hearing aid use (Fraser, Kenyon, Lagacé, Wittich, & Southall, 2015; Gagné et al., 2009; Kochkin, 2007; Southall et al., 2010). Hearing impairment has an important impact on quality of life. Not only does it bring its share of functional and communication difficulties (Mulrow et al., 1990); it is also associated with stigma that can create important social and emotional hardships. The stigma associated with hearing loss often incorporates ageist stereotypes (Coleman, 2006; Espmark & Scherman, 2003; Fraser et al., 2015; Southall et al., 2010; Tannenbaum et al., 2015). It is common for older adults with hearing loss to quickly give up on their hearing aids or simply refuse to use them due to fear of stigmatization. Some of them eventually stop participating in social activities altogether because they do not want to be perceived as being "deaf" or to be seen with hearing aids (Kochkin, 2007). Accordingly, interventions that involve working on the negative perceptions of hearing loss with this population have been proposed in order to encourage people with hearing impairment to seek rehabilitation services (Hetu, 1996).

When implementing a new intervention program designed for people with hearing loss who self-stigmatize, it may be useful to appraise the client's perception of the stigmatizing trait before, during, and after the program. Unfortunately, only English-language measures are available for individuals of other stigmatized groups such as women, gay men and lesbians, ethnic communities (Lewis, Derlega, Griffin, & Krowinski, 2003; Pinel, 1999), and people diagnosed with mental illness (King et al., 2007). In audiology, there is a need to measure stigma associated with hearing loss, and to do so it must be done in the mother tongue of the client. Quebec's population (7,651,000) accounts for 23.9% of the Canadian population, and Quebec's francophones account for at least 90% of all of Canada's French-speaking population (Marmen & Corbeil, 2004). The importance of measuring stigma includes the need for researchers and professionals in all bilingual regions to have access to valid and reliable instruments in both French and English.

This article presents the initial development of transcultural validation of two questionnaires assessing different aspects of stigma and its consequences among older adults with hearing loss. This article also serves as a model for the rigorous process that may be used to translate and adapt existing measurement tools in another language. Moreover, the psychometric properties of the original English versions and the adapted Canadian-French versions of the questionnaires are compared in this article.

Description of the Stigma Consciousness Questionnaire (SCQ)

The SCQ is a 10-item self-report questionnaire that measures the extent to which respondents expect to be stereotyped because of their disability, social role, or sexual orientation. It is also intended to measure how this affects the way respondents experience their stereotyped status (Pinel, 1999). This questionnaire was initially developed to measure stigma associated with being a woman (development $N = 722$ and final form tested on $N = 302$; Pinel, 1999). In subsequent studies, Pinel (1999) tested the generalizability of the stigma-consciousness construct by adapting and validating the scale for gay men ($n = 23$) and lesbians ($n = 27$), Caucasians ($n = 198$), Asians ($n = 63$), Hispanics ($n = 53$), and Afro-Americans ($n = 21$).

The initial version of the SCQ focused on two domains: (1) the phenomenological experiences of women when interacting with men (e.g., "I never worry that my behaviors will be viewed as typically female") and (2) beliefs on how men view women (e.g., "Most men have a lot more sexist thoughts than they actually express"). The scale questioned women about their perceptions of how they are judged by men and of how differently men interact with them.

When answering the SCQ for Women, respondents are asked to read each of the 10 statements and indicate to what extent they agree with each statement by rating them on a 7-point scale ranging from 0 (completely disagree) to 6 (completely agree). The scale includes a midpoint of 3, denoting "neither agree nor disagree". Seven of the 10 items are reverse scored. A high total score indicates that a respondent's level of stigma consciousness is high. In other words, the respondent is strongly concerned with how others view him or her and is more aware of the signs of sexism. The evaluation of the instrument's internal consistency as well as the discriminant and convergent validities were evaluated by comparing the SCQ to other instruments assessing concepts such as self-consciousness, modern sexism, and gender attitudes. The instrument's construct validity and evaluation of test-retest reliability were also performed (Pinel, 1999).

Description of the Stigma Scale for Mental Illness

The Stigma Scale for Mental Illness is a 28-item instrument that asks respondents about their experiences of discrimination and their feelings concerning prejudice. It is divided into three subscales: discrimination, disclosure, and positive aspects (King et al., 2007). The first subscale is composed of 13 items that address the discriminatory attitudes of others and their consequences (e.g., lost opportunities) as perceived by the respondent. The second subscale is composed of 10 items that address the respondent's embarrassment concerning mental illness and his or her way of managing disclosure in order to avoid discrimination. The third subscale is composed of five items that question the respondent's perspective on the possibility that having a mental illness has made him or her a better person (e.g., more understanding and accepting of others).

When this stigma scale is administered, respondents are asked to read the 28 statements and check off the answer that best corresponds to each of the statements. Respondents are asked not to ponder too long on each question because the questionnaire aims to obtain their first impression. Response options vary from "strongly agree" to "strongly disagree" on a 5-point Likert-type scale. A higher score on the stigma scale for mental illness is indicative of a greater amount of stigma. Evaluation of the instrument's test-retest reliability and internal consistency was performed (King et al., 2007).

Methods

Instrument translation

For both questionnaires assessing hearing loss stigma, the translation protocol used was inspired by the initial steps of the methodology proposed by Vallerand (1989). As outlined in Figure 1, the first step of the procedure involved preparing preliminary versions of the original questionnaires. The research team began by confirming that each original instrument was correctly adapted to the phenomenon of stigmatization generated by a hearing disability. Both questionnaires were modified since they originally targeted groups other than persons with hearing impairment. Consequently, the SCQ was adapted for this clientele and identified as the Canadian-French Stigma Consciousness Questionnaire for Hearing Loss (SCQ-CF). In French, the scale is referred to as the *Échelle de la conscience de la stigmatisation personnelle (associée à la perte d'audition).*

Similarly, the Stigma Scale for Mental Illness was adapted to target persons with hearing loss and was identified as the Canadian-French Hearing Loss Stigma (HLS) questionnaire (HLS-CF). In French, this questionnaire is entitled *Stigmatisation associée à la déficience auditive.*

Two other translations of each questionnaire were then performed independently, one by a member of the research team and the other by a professional translator with no particular background in the health domain. French was the native language of both individuals who translated the questionnaires. Then, using solely the French versions, a back-translation into English was done independently by two experienced Anglophone audiologists blinded to the original English questionnaires.

Following that step, a committee of experts ($N = 4$) consisting of the research team and the professional translator met in order to review the preliminary French versions of each questionnaire and to generate only one French experimental version. The back-translated English versions were compared to their respective original (English) versions to see how much the original and translated versions were alike. The more closely the back-translated versions resembled the original English versions, the more they were deemed accurate. When discrepancies in wording were observed, the committee examined both translated versions carefully and decided which wording was the most accurate. The same process was used to translate the titles, the instruments' introductory text, the instructions, and the response options in order to obtain satisfactory experimental versions formatted similarly to the original instruments.

Experimental versions

Ethical approval was obtained from the Research Ethics Board of the Centre de recherche interdisciplinaire en réadaptation du Montréal métropolitain (CRIR-731-0412). Before taking part in the study, each participant read and signed an informed consent form. The experimental versions of the questionnaires were administered to ensure that there was no ambiguity and to assess the validity of the content. Both the translated and original versions were administered to five bilingual persons using a two-step procedure. The questionnaires (see Appendices A–D) were administered in a pre-established order (SCQ French version, SCQ English version, then HLS English version and HLS French version). In the methodology proposed by Vallerand (1989), five participants are recommended.

The first part was carried out with a bilingual 79-year-old retired woman without hearing loss. Her native language was Canadian French and she learned English while working as a secretary at an anglophone accounting firm. With the help of a research team member, this participant

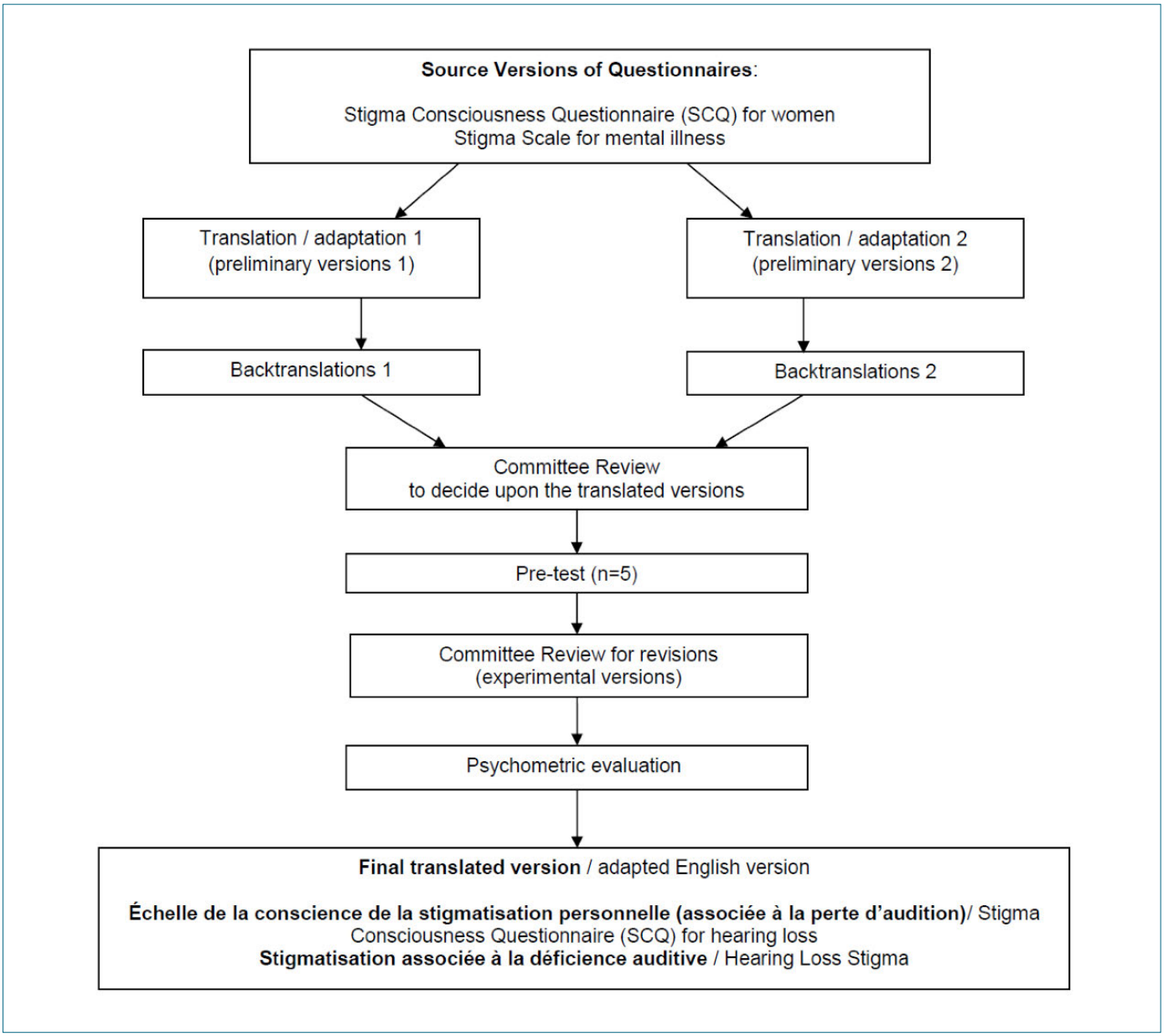

Figure 1. Adaptation and translation protocol for both stigma questionnaires

completed both versions of each questionnaire. For each questionnaire, she was asked to point out any ambiguities between the French and the English versions of the same test items, as well as the introductory text and test instructions of the translated versions. She was also asked to identify any differences in meaning that she noticed between both versions of the questionnaires. Following the administration of the questionnaire, the research team member compared the answers and collected the participant's comments. Whenever a mismatch occurred in the answers provided for the same test item in the two languages, the research team member discussed the nature of the ambiguity with the participant. Whenever an item was unclear, the following rule was applied: If the ambiguous item only occurred in the French version, the unclear item was rephrased as needed. If, on the other hand, the same item was considered unclear in both languages, it was left unchanged. Once this step was completed, the research team and the professional translator met to validate the revisions made to the translated versions of the two instruments.

The second part of the administration of the experimental versions involved the participation of three females and one male. The four participants were between 66 and 82 years of age and all of them had hearing loss

ranging from mild to severe. This was an important aspect of the experimental versions because it was the first time that the questionnaires were administered to participants who had the same profile as the intended population (i.e., older adults with hearing loss). One participant was a retired office manager, and another was a retired financial advisor. The third person was a designer and the fourth was a translator. The participants were asked to complete the French and English versions of both questionnaires and point out areas of ambiguity. The purpose of this step was to further improve the translated experimental versions.

Evaluation of the psychometric properties of the translated versions

According to Vallerand's methodology (1989), the research team needed to complete the transcultural validation by reproducing and reporting the same tests as the original (English) questionnaire. This had been done for the SCQ-CF and the HLS-CF versions for total scores, internal consistency, test-retest stability and factor analysis. The mean total score and standard deviation are presented for both versions as well as for the sub-scores when available. Calculations were made with G*Power software, version 3.1.7, using sample sizes, means and standard variations to evaluate significant differences with the original article.

Statistics were calculated using SPSS 23 software. If missing data were encountered, we used the mean answer from two other participants having the most similar answer to the participant with the missing value (based on hot-deck imputation). The underlying principle was that researchers were to replace a missing value with the actual score from a similar case in the current data set (Roth, 1994). To replace each missing data point in the present study, the two participants with the most similar response patterns were identified from the 31 other participants, and the mean of their answers was used to fill in the missing answer. Internal consistency was assessed to examine the degree to which the items that made up each scale were homogenous. The coefficient of reliability computed was Cronbach's alpha (α), which can range between 0 and 1. In a good questionnaire, items must be balanced between homogeneity and diversity; this is why some authors suggest using a range between 0.7 and 0.9 (Boyle, 1991; Hyde, 2000; Norman & Streiner, 1999). Temporal stability was assessed by administering each translated questionnaire on two separate occasions in order to ascertain the correlation between the two sets of scores. The second test session took place approximately four weeks after initial testing. In psychometrics literature, an interval of 2 days to 6 weeks is acceptable (Cohen, Kamarck, & Mermelstein, 1983; Lee, 2012). A two-week interval is generally acceptable to minimize the carryover effects due to memory and to limit the possibility of a change in participant status (Marx, Menezes, Horovitz, Jones, & Warren, 2003). Finally, for each translated questionnaire, an alpha-maximized factor analysis (oblimin rotation, delta = 0) was conducted to assess the underlying latent variables. This psychometric property verifies whether the items of a scale cluster into the appropriate subscales, as supported by the theory.

With our expected sample size we knew that the factor analysis for the HLS would not be conclusive ($N = 32$ for 28 items), but we reported the data to show the results that would be obtained with this questionnaire using this sample size. Otherwise, it was postulated that the psychometric properties of the translated questionnaires could be at least as good as those of the original test versions.

Participants

A convenient sample of 32 persons was proposed based on the feasibility of a stigma group intervention program. By "convenient" we mean that we recruited persons who initially wanted to participate in a research project focusing on preventing stigma related to hearing loss and wearing hearing aids. The intervention program consisted of two groups of participants ($14 > n < 16$) that met together to practice adaptive strategies and conduct debriefings with peers. We took this opportunity to ask participants if they would be willing to validate the questionnaires for the study in French and they all agreed. Older adults were recruited with the help of audiologists from the *Centre intégré universitaire de santé et de services sociaux (CIUSSS) du Centre-sud de l'île de Montréal - Institut Raymond-Dewar* (a rehabilitation centre specializing in services for persons with hearing impairment). The audiologists informed their patients that a research project on hearing difficulties was taking place. Patients who wanted to participate in the study were invited to contact a member of the research team. Participant inclusion criteria were: (1) to be willing to participate and (2) to recognize having some hearing difficulties or issues associated with hearing loss. Hearing aid ownership as well as a clinical diagnosis of hearing loss were not required to participate. Finally, because of their unique profile, cochlear implant users were excluded from the study. Recruited patients signed a consent form.

Results

The individuals selected for the study were adults who were 65 years of age or older ($N = 32$). The majority of them ($n = 24$) were hearing aid owners. Table 1 provides a summary description of the participants involved in the validation of the questionnaires.

Table 1. Characteristics of Participants Involved in Canadian-French Validation of the Stigma Consciousness Questionnaire (SCQ) and the Hearing Loss Stigma (HLS) Questionnaire

Degree of hearing loss	Sample size *(N = 32)*	Hearing aid users	Demographics		Psychometric properties investigated
			Mean age (range)	Male:Female ratio	
Mild (15-40 dB)	7	28%	73.0 (68-80)	2:5	
Moderate (41-70 dB)	19	84%	75.5 (66-90)	5:14	Internal consistency Factor analysis Temporal stability
Severe (71+ dB)	6	100%	78.2 (70-86)	1:5	

Final Canadian-French version and comparison of psychometric properties of the SCQ

The translated versions of the questionnaire appear in Table 2. Every test item from the original English questionnaire is presented, along with the accompanying tests items from the translated Canadian-French version of the questionnaire.

Psychometric properties of the SCQ-CF as well as the psychometric properties of the respective original English questionnaire are displayed in Table 3. The latter are placed immediately under the results of the translated version to facilitate comparison.

Table 2. Item Translation of the Stigma Consciousness Questionnaire[1] (SCQ) for Hearing Loss

Item	English version	Canadian-French (CF) translation
1	Stereotypes about hearing loss have not affected me personally.	Les stéréotypes concernant la perte d'audition ne m'affectent pas personnellement.
2	I never worry that my behaviors will be viewed as stereotypical of a person who has a hearing loss.	Je ne suis jamais inquiet que mes comportements puissent être perçus comme étant typiques d'une personne ayant une perte d'audition.
3	When interacting with people who have normal hearing, I feel like they interpret all my behaviors in terms of the fact that I have a hearing loss.	Lorsque je dialogue avec des gens ayant une audition normale, je sens qu'ils interprètent tous mes comportements en fonction du fait que j'ai une perte d'audition.
4	Most people with normal hearing do not judge people with hearing loss on the basis of their ability to hear.	La plupart des gens ayant une audition normale ne jugent pas les gens ayant une perte d'audition en fonction de leur capacité à entendre.
5	My being hearing impaired does not influence how people with normal hearing act with me.	Le fait que je suis une personne malentendante n'a aucune influence sur la façon dont les gens ayant une audition normale agissent avec moi.
6	I almost never think about the fact that I have a hearing loss when I interact with people who have normal hearing.	Je ne pense presque jamais au fait que j'ai une perte d'audition lorsque je dialogue avec quelqu'un qui a une audition normale.
7	My being hearing impaired does not influence how people act with me.	Le fait que je suis une personne malentendante n'a aucune influence sur la façon dont les gens agissent avec moi.
8	Most people with normal hearing have a lot more prejudicial thoughts about people with hearing loss than they actually express.	La plupart des gens ayant une audition normale ont beaucoup plus de préjugés à l'endroit des personnes ayant une perte d'audition qu'ils ne le disent en réalité.
9*	I often think that people with normal hearing are unfairly accused of having prejudicial thoughts about people with hearing loss.*	Je crois souvent que les personnes ayant une audition normale sont injustement accusées d'avoir des préjugés envers les personnes qui ont une perte d'audition.*
10	Most people with normal hearing have a problem viewing people with hearing loss as equals.	La plupart des gens qui ont une audition normale ont de la difficulté à considérer les personnes qui ont une perte d'audition comme étant des égaux.

Note. [1]Translated as *Échelle de la conscience de la stigmatisation personnelle (associée à la perte d'audition).* "SCQ-CF for Hearing Loss" is a better option to keep the original questionnaire in mind. *The temporal reproducibility of this item (9) was not observed. Do not take this into account during longitudinal follow-up.

Table 3. Comparison Between the Psychometric Properties of the Original Stigma Consciousness Questionnaire (SCQ) for Women[1]

		French	English
	N =	32	201[2]
Total score (SD)		37.1 (9.7)*	23.6 (6.8)
Factor analysis		3 factors	1 factor
Eigenvalues		3.9, 1.5 and 1.3	-
Total variance explained		67% (39% + 15% + 13%)	23%
Common variance explained		55% (35% + 11% + 9%)	91%
KMO		.704	-
Bartlett's test		< .001	-
Determinant		0.010	-
	N =	32	302[1]
Internal Consistency (α)		.79	.72
No. item increases α if deleted (new α)		2 (0.80)	0
	N =	32	57
Temporal Stability (ICC): *r*		.62	.76
Number of non-significant items (#)		1 (#9)	0
Time interval (weeks)		4	5

Notes. [1]From Pinel, 1999, study 1; [2]From Pinel, 1999, study 5; *Statistically different from the English version (α probability of error < .05). Total score was made by an addition of the items after the inversion of the score of the appropriate items (1, 2, 4, 5, 6, 7 and 9); SD = Standard deviation; KMO = Kaiser-Meyer-Olkin, degree of collinearity between variables; ICC = Intraclass correlation coefficient.

When comparing the Canadian-French and English versions, it can be observed that the latent variables of the SCQ-CF are different from the original English version. The alpha-maximized factor analysis revealed three factors after an oblimin rotation (eigenvalues: 3.9, 1.5 and 1.3). These three factors account for 67% of the total variance (39% + 15% + 13%). The determinant was 0.010, indicating an absence of multicollinearity. The Kaiser-Meyer-Olkin (KMO) measure, indicating the degree of collinearity between variables, was satisfying (.704) and Bartlett's test was significant (< .001), demonstrating the absence of an identity matrix.

In comparison, the original article identified only one factor (principal-axis factor analysis), which accounts for 23% of the total variance. Of the three factors identified in the present study, the first one contains items 4 to 7. These items are related to interaction with others, especially people with normal hearing. The second factor contains items 1, 2, and 9. These statements are associated with the feeling respondents have about the stereotypes they project onto others. The third factor contains items 3, 8, and 10. These variables are related to respondents' perceived truthfulness of normal hearing people.

The Canadian-French version of the SCQ for hearing loss has a Cronbach α value of .79. The removal of two items slightly increased the α value to .80. No other change in the α value was observed when other test items were removed (lowest: .74). Cronbach's α was also evaluated for the three identified factors (Factor 1: .88, Factor 2: .62, and Factor 3: .56). For the second factor, the removal of item 2 reduced the α value to .31, indicating that it was the main constituent of this factor. The same is true for the third factor, where the removal of item 8 induced a marked decrease in α to .38. In comparison, the original article reported a lower α value of .72 (no item increase the α when an item is removed).

The Canadian-French SCQ for hearing loss had a test-retest correlation of .62 for the global score, compared to .76 in the original article. The time elapsed between testing sessions was 4 weeks for our study, compared to a mean of 5 weeks in the original article. The correlations were significant for nine of the 10 test items, and the correlation score ranged from .38 to .66. We did not find any correlation for item 9 (p = .20; r = .15).

Final Canadian-French version and comparison of psychometric properties of the HLS

Every test item of the original English HLS questionnaire is presented in Table 4, along with the accompanying test items of the translated Canadian-French version of the questionnaire.

Psychometric properties of the Canadian-French version of the HLS as well as the psychometric properties of the original English questionnaire are displayed in Table 5.

For this questionnaire, there were 12 missing data points in total, out of all 32 participants answering 28 questions each, representing only 1.3% of all answers. The imputation method has been previously described in the Methods section. The factor analysis could not be used due to the small number of participants (N = 32 for 28 items), as indicated by the low KMO (.415). In the original article, the investigators recruited an extra 100 participants in order to be able to perform this analysis. The internal consistency of the HLS-CF shows an equivalent coefficient compared to the original English version of the questionnaire (.91 versus .87, respectively). The removal of items did not increase the alpha value significantly. The internal consistency was also calculated for the three sub-scores. The discrimination and disclosure subscales showed high and comparable alpha values (.91 versus .87, respectively) compared to the original article (.87 and .85). We observed a much lower alpha value for the positive aspect subscale (.33). A lower internal consistency was also observed in the original article for this subscale.

The time elapsed between testing sessions was 4 weeks for our study, compared to a mean of 2 weeks in the original article. The original article reported that the coefficient of correlations ranged from .40 to .71 for individual items. The same range of individual correlations was observed for 25 of the 28 items (.42-.74). For the other three test items, the correlation was not significant (items 5, 6, and 15; r = .05-.28).

Discussion

The goal of this study was to adapt, translate, and validate two questionnaires that assess different aspects of stigma and its consequences among older adults with hearing loss. We hypothesized that the psychometric properties would be as good as in the original article, even if their sample sizes were made of other clienteles with a potentially lower chance of exposure to loud noise. We obtained very good comparisons for internal consistency and temporal stability of the HLS-CF, even if the factor analysis was not conclusive as anticipated. The SCQ-CF showed better internal consistency than the original article but a lower temporal stability and different results for the factor analysis. With these results, we did not reach an overall validation for the novel questionnaires, but these initial developments constitute a respectable preliminary validation. These aspects will be analyzed in detail in this section.

Canadian-French Stigma Consciousness Questionnaire (SCQ) for Hearing Loss

Even though the participant/item ratio is low (3.2/1), we performed an alpha-max factor analysis that indicated a satisfactory determinant, KMO and Bartlett's test. This factor analysis maximizes Cronbach's α for each factor. The oblimin rotation allows the factors not to be orthogonal, and in the case where there is truly no correlation between factors, the results are the same as those for a varimax. A correlation matrix between factors after the oblimin rotation shows correlation between factors 1 and 2 (-.31) and between factors 1 and 3 (.27), but no correlation between factors 2 and 3 (-.04). Tabachnick and Fidell (2007, p. 646) indicate that a correlation under .32 is considered orthogonal because the factors have less than 10% common variance. That we identified three factors rather than only one may be due to the fact that the questionnaire was adapted to a new population that perceives stereotypes differently. We are not the only

Table 4. Item Translation of the Hearing Loss Stigma[1] (HLS) Questionnaire

Item	English version	Canadian-French (CF) translation
1	I have been discriminated against in education because of my hearing problems.	J'ai subi la discrimination durant mes études en raison de mes problèmes d'audition.
2	Sometimes I feel that I am being talked down to because of my hearing problems.	J'ai parfois l'impression qu'on me rabaisse en raison de mes problèmes d'audition.
3	Having had hearing problems has made me a more understanding person.	Mes problèmes d'audition ont fait de moi une personne plus compréhensive.
4	I do not feel bad about having hearing problems.	Je ne m'en fais pas à propos de mes problèmes d'audition.
5*	I worry about telling people I received help concerning my hearing problems.*	Je crains de dire aux gens que j'ai reçu de l'aide pour mes problèmes d'audition.*
6*	Some people with hearing problems are cognitively challenged.*	Certaines personnes ayant des problèmes d'audition ont des difficultés cognitives.*
7	People have been understanding of my hearing loss.	Les gens se montrent compréhensifs à l'égard de mes problèmes d'audition.
8	I have been discriminated against by friends and relatives because of my hearing problems.	Je subis la discrimination de la part de mes amis et parents en raison de mes problèmes d'audition.
9	I have been discriminated against by employers because of my hearing problems.	Je subis la discrimination de la part d'employeurs en raison de mes problèmes d'audition.
10	My hearing problems have made me more accepting of other people.	Mes problèmes d'audition ont fait de moi une personne qui accepte mieux les autres.
11	Very often I feel alone because of my hearing problems.	Il m'arrive très souvent de me sentir seul(e) en raison de mes problèmes d'audition.
12	I am scared of how other people will react if they find out about my hearing problems.	Je crains la façon dont les autres personnes réagiront si elles découvrent mes problèmes d'audition.
13	I would have had a better chance in life if I did not have hearing problems.	J'aurais eu plus de chance dans la vie si je n'avais pas eu de problèmes d'audition.
14	I do not mind people in my neighborhood knowing I have hearing problems.	Cela ne me dérange pas que les gens de mon voisinage soient au courant de mes problèmes d'audition.

15*	I would say I have hearing problems if I was applying for a job.*	Je dévoilerais mes problèmes d'audition si je postulais un emploi.*
16	I worry about telling people that I have consulted hearing experts about my hearing problems.	Je crains de dire aux gens que j'ai consultés des spécialistes pour mes problèmes d'audition.
17	People's reactions to my hearing problems make me keep myself to myself.	Les réactions des gens à mes problèmes d'audition m'amènent à ne pas me livrer.
18	I am angry with the way people have reacted to my hearing problems.	La façon dont les gens réagissent à mes problèmes d'audition me met en colère.
19	I have not had any trouble from people because of my hearing problems.	Les gens ne m'ont jamais causé d'ennuis en raison de mes problèmes d'audition.
20	I have been discriminated against by health professionals because of my hearing problems.	Je subis la discrimination de la part de professionnels de la santé en raison de mes problèmes d'audition.
21	People have avoided me because of my hearing problems.	Des gens m'ont évité(e) en raison de mes problèmes d'audition.
22	People have insulted me because of my hearing problems.	Des gens m'ont insulté(e) en raison de mes problèmes d'audition.
23	Having hearing problems has made me a stronger person.	Mes problèmes d'audition ont fait de moi une personne plus forte.
24	I do not feel embarrassed because of my hearing problems.	Je ne suis pas gêné(e) de mes problèmes d'audition.
25	I avoid telling people about my hearing problems.	J'évite de dévoiler mes problèmes d'audition aux gens.
26	Having hearing problems makes me feel that life is unfair.	Mes problèmes d'audition m'amènent à penser que la vie est injuste.
27	I feel the need to hide my hearing problems from my friends.	Je sens le besoin de cacher mes problèmes d'audition à mes amis.
28	I find it hard telling people I have hearing problems.	J'ai de la difficulté à dévoiler mes problèmes d'audition aux gens.

Note. [1]Translated as *Stigmatisation associée à la déficience auditive.* HLS-CF is a better option to keep the original questionnaire in mind. *The temporal reproducibility of these items (5, 6 and 15) was not observed. Do not take this into account during longitudinal follow-up.

Table 5. Comparison Between the Psychometric Properties of the Original Stigma Scale for Mental Illness[1]

		French	English
	N =	32	185-192
Mean scores[2] (SD)			
Total		33.7 (16.3) *	62.6 (15.4)
Discrimination		15.8 (10.6) *	29.1 (9.5)
Disclosure		10.0 (6.6) *	24.7 (8.0)
Positive aspect		7.9 (2.6)	8.8 (2.8)
	N =	32	163
Factor analysis		Not Valid	3 factors
Eigenvalues		NA	7.7, 2.8 and 2.1
Total variance explained		NA	72% (44% + 16% +12%)
KMO		.415	-
Bartlett's test		< .001	-
	N =	32	93
Internal Consistency (α)			
All items		.91	.88
Discrimination		.91	.87
Disclosure		.87	.85
Positive aspect		.33	.64
	N =	32	60
Temporal Stability [3] (ICC) : *r*		.42-.73	.40-.71
Number of non-significant items (#)		3 (5, 6 and 15)	0
Time interval (weeks)		4	2

Notes. [1]From King et al., 2007; [2]Each question scored 0-4 in the direction of greater stigma. Total score and subscores were made by a an addition of the items; [3]Presented item by item in the original article; *Statistically different from the English version (α probability of error < .05); SD = Standard deviation; KMO = Kaiser-Meyer-Olkin, degree of collinearity between variables; ICC = Intraclass correlation coefficient.

authors to have observed these three factors for a different version of the SCQ. An academic work by Huie shows the same three factors, composed of the same items, after a factor analysis performed on 149 participants who completed the Stigma Consciousness Questionnaire for Race (Huie, 2010). This cannot be a coincidence, and might explain the fact that the relationship between the normal-hearing population and the hearing-impaired population (or population with and without mental illness, in the case of Huie's work) differs compared to the relationship between women and men. We also need to consider that factors 2 and 3 are mainly driven by items 2 and 8 respectively, and in Pinel (1999), the factor loading for items 2 and 8 was among the weakest associated with the unique factor (.33 and .40). For this questionnaire, we obtained a lower test-retest score compared to the original article for the global score (.62 versus .76) and we observed that nine of the 10 items had significant test-retest correlations, ranging from .38 to .66. It is thus unlikely that the removal of the non-significant item (item 9) would increase the global test-retest score to the level of the original article. However, by looking more closely at this item, we realized that seven participants switched their answers to this question during the test-retest (i.e., from agree to disagree or vice versa). By replacing only one of these drastic changes (e.g., completely agree to completely disagree), we obtained a significant correlation (p = .036; r = .373). We are not sure what could have motivated these drastic changes of opinion for this particular item. Further investigations could have be done to verify whether participants understood this item, especially when item 9 was also one of the two items that increased Cronbach's α very slightly (from .79 to .80) when it was removed.[1] However, this increase is very small and since we did not observe large variations in Cronbach's α when items were removed, we believe this questionnaire is reliable for measuring stigma consciousness. In 1999, Pinel reported total scores for different populations and versions of this questionnaire. We report a total score that is significantly higher than the one reported for women but is quite similar to the one calculated for men (n = 142, mean 33.5, SD 6.83; α probability of error .29).

Canadian-French Hearing Loss Stigma (HLS) questionnaire

The total score for our questionnaire is significantly lower than the one reported for mental illness stigma. Scores for the discrimination and disclosure subscales are also significantly lower, but the positive aspect subscale is not different. We were not surprised to see these differences, considering the different populations studied. The factor analysis was also not considered due to the low KMO and lack of participants.

An excellent and comparable Cronbach α value was obtained for the adapted instrument compared to the one reported for the original questionnaire (.91 versus .88). The internal consistency is also similar for the discrimination and disclosure subscales. Cronbach's α for the positive aspect is very low even when we consider that in the original version it was by far the subscale with the lowest Cronbach α value. The removal of item 6 greatly increased the alpha value (to .55), indicating that it no longer fits in this subscale. The meaning of this item has been adapted between questionnaires, and it is normal to observe this difference (going from "Some people with mental health problems are dangerous" to "Some people with hearing problems are cognitively challenged").

We observed a comparable range of correlation coefficients for the 25 items showing significant correlation. Only three items (5, 6, and 15) had no correlation in the test-retest (see Footnote 1). A closer examination of the data revealed that the problem did not come from the fact that answers were random on the retest. Again, the problem comes from the fact that two participants had chosen opposite ratings relative to their original answers. The replacement of only one of these answers makes the correlations significant. These two participants had no other surprising answers and the two problematic answers were not part of the imputed data. We could also verify the interpretation of these three items to determine if the participants' understanding of the question differed.

Limits of the study and future research

An increase in sample size would help to confirm the underlying latent variables of the questionnaires. In addition, validation measures targeting stigma need to be conducted on two additional questionnaires that we have prepared for use with Canadian-French-speaking adults: the International Outcome Inventory - Alternative Interventions (IOI-AI) (Laplante-Lévesque, Hickson, & Worrall, 2012) and the Expected Consequences of Hearing aid Ownership (ECHO) (Cox & Alexander, 2000). Since this is a first step in producing two questionnaires in French, it should be seen as a preliminary validation, and in that sense, more research is needed to enhance the psychometric qualities of those novel questionnaires. Additional validation work is needed to clarify some of the issues that arose during this first transcultural study (e.g., reproducibility of some items,

[1]In both versions of this questionnaire, an asterisk (*) has been added with a footnote to invite the clinicians not to include poorly reproducible items when computing subscores in longitudinal follow-ups.

missing data, item 9 on the SCQ, factorial analysis, and small sample size ($N = 32$)).

Conclusion

This transcultural validation study successfully adapted two questionnaires that address the concept of stigma associated with hearing loss. Results indicate that they both have psychometric properties comparable to the versions used for the stigma against women and mental illness, with the exception of the latent variables, since they have been slightly modified for the hearing impaired. Readers who would like to obtain the questionnaires produced in this study should contact one of the first two authors of this manuscript. These two novel questionnaires addressing stigma (SCQ and HLS) provide useful information for clinicians when working with French- and English-speaking adults. Clinical implementation of these questionnaires should be incorporated into practice. Since it is more important to address stigma in clinics than to ignore it, we encourage clinicians to use both questionnaires in French and in English but to be careful with the interpretation of items having poor or weak reproducibility. For all items where reproducibility is still a challenge (only four), an asterisk (*) has been added with a footnote to invite the clinicians not to include them when computing subscores. Also, French and English clinicians are welcome to email to authors with any suggestions to upgrade item formulation. These novel questionnaires could also be used while larger validation studies are being undertaken, especially with respect to factor analysis, since this was the first time these questionnaires have been used for older adults with hearing loss.

References

Boyle, C. M. (1991). Does item homogeneity indicate internal consistency or item redundancy in psychometric scales? *Personality and Individual Differences, 12*(3), 291–294. doi: 10.1016/0191-8869(91)90115-R

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior, 24*(4), 385–396.

Coleman, L. M. (2006). Stigma: An enigma demystified. In L. Davis (Ed.), *The disability studies reader* (2nd ed., pp. 141-152). New York, NY: Routledge.

Cox, R. M., & Alexander, G. C. (2000). Expectations about hearing aids and their relationship to fitting outcome. *Journal of the American Academy of Audiology, 11*(7), 368–382.

Espmark, A. K. K., & Scherman, M. H. (2003). Hearing confirms existence and identity: Experiences from persons with presbyacusis. *International Journal of Audiology, 42*(2), 106–115. doi: 10.3109/14992020309078341

Fraser, S.A., Kenyon, V., Lagacé, M., Wittich, W., & Southall, K. E. (2016). Stereotypes associated with age-related conditions and assistive device use in Canadian media. *Gerontologist, 56*(6): 1023–1032. doi: 10.1093/geront/gnv094

Gagné, J.-P., Southall, K., & Jennings, M. B. (2009). The psychological effects of social stigma: Applications to people with an acquired hearing loss. In J. Montano & J. Spitzer (Eds.), *Advanced practice in adult audiologic rehabilitation: International perspective* (pp. 63–92). San Diego, CA: Plural Publishing Inc.

Heine, C., & Browning, C. J. (2002). Communication and psychosocial consequences of sensory loss in older adults: Overview and rehabilitation directions. *Disability and Rehabilitation, 24*(15), 763–773. doi: 10.1080/09638280210129162

Hétu, R. (1996). The stigma attached to hearing impairment. *Scandinavian Audiology Supplementum, 43*, 12–24.

Huie, F. (2010). *Stigma consciousness and academic achievement: The role of self-regulatory processes.* Unpublished manuscript, Graduate School of Education, George Mason University, Fairfax, VA. Retrieved from mason.gmu.edu/~fhuie/portfolio/AdvancedQuantFinalPaper.doc

Hyde, M. L. (2000). Reasonable psychometric standards for self-report outcome measures in audiologic rehabilitation. *Ear Hear, 21*, 24S–36S.

Jennings, M. B., Southall, K., & Gagné, J. P. (2013). Social identity management strategies used by workers with acquired hearing loss. *Work, 46*(2), 169–180. doi: 10.3233/WOR-131760

King, M., Dinos, S., Shaw, J., Watson, R., Stevens, S., Passetti, F., & Serfaty, M. (2007). The Stigma Scale: Development of a standardised measure of the stigma of mental illness. *British Journal of Psychiatry, 190*, 248–254. doi: 10.1192/bjp.bp.106.024638

Kochkin, S. (2007). MarkeTrak VII: Obstacles to adult non-user adoption of hearing aids. *Hearing Journal, 60*(4), 24–51. doi: 10.1097/01.HJ.0000285745.08599.7f

Laplante-Lévesque, A., Hickson, L., & Worrall, L. (2012). Comparing response options for the International Outcome Inventory for Hearing Aids (IOI-HA) and for Alternative Interventions (IOI-AI) daily-use items. *International Journal of Audiology, 51*(10), 788–791. doi: 10.3109/14992027.2012.695875

Lee, E. H. (2012). Review of the psychometric evidence of the Perceived Stress Scale. *Asian Nursing Research, 6*(4), 121–127. doi: 10.1016/j.anr.2012.08.004

Lewis, R. J., Derlega, V. J., Griffin, J. L., & Krowinski, A. C. (2003). Stressors for gay men and lesbians: Life stress, gay-related stress, stigma consciousness, and depressive symptoms. *Journal of Social and Clinical Psychology, 22*(6), 716–729. doi: 10.1521/jscp.22.6.716.22932

Marmen, L., & Corbeil, J. P. (2004). *New Canadian perspectives. Languages in Canada 2001 census* (Statistics Canada Cat. No. CH3-2/8-2004). Retrieved from http://publications.gc.ca/collections/Collection/CH3-2-8-2004E.pdf

Marx, R. G., Menezes, A., Horovitz. L., Jones. E. C., & Warren, R. F. (2003). A comparison of two time intervals for test-retest reliability of health status instruments. *Journal of Clinical Epidemiology, 56*(8), 730–735. doi: 10.1016/S0895-4356(03)00084-2

Mulrow, C. D., Aguilar C., Endicott, J. E., Tuley, M. R., Velez, R., Charlip, W. S., ... DeNino, L. A. (1990). Quality-of-life changes and hearing impairment: A randomnized trial. *Annals of Internal Medicine, 113*(3), 188–194. doi: 10.7326/0003-4819-113-3-188

Parrette, P. & Scherer, M. (2004). Assistive technology use and stigma. *Education and Training in Developmental Disabilities, 39*(3), 217–226.

Pinel, E. C. (1999). Stigma consciousness: The psychological legacy of social stereotypes. *Journal of Personality and Social Psychology, 76*(1), 114–128. doi: 10.1037/0022-3514.76.1.114

Roth, P. L. (1994). Missing data: A conceptual review for applied psychologists. *Personnel Psychology, 47*(3), 537–560. doi: 10.1111/j.1744-6570.1994.tb01736.x

Southall, K., Gagné, J.-P., & Jennings, M. B. (2010). Stigma: A negative and a positive influence on help-seeking for adults with acquired hearing loss. *International Journal of Audiology, 49*(11), 804–814. doi: 10.3109/14992027.2010.498447

Streiner, D. L., & Norman, G. R. (1995). *Health measurement scales. A practical guide to their development and use.* New York, NY: Oxford University Press.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Upper Saddle River, NJ: Pearson/Allyn & Bacon.

Tannenbaum, C., van den Heuvel, E., Fritel, X., Southall, K., Jutai, J., Rajabali, S., & Wagg, A. (2015). Continence Across Continents To Upend Stigma and Dependency (CACTUS-D): Study protocol for a cluster randomized controlled trial. *Trials, 16*, 1-11. doi: 10.1186/s13063-015-1099-x

Vallerand, R. J. (1989). Vers une méthodologie de validation trans-culturelle de questionnaires psychologiques: implications pour la recherche en langue française. *Canadian Psychology, 30*(4), 662–680. doi: 10.1037/h0079856

Acknowledgements

This study was made possible by funding from the Partnership of the Réseau provincial de recherche en adaptation-réadaptation (REPAR), the Centre intégré universitaire de santé et des services sociaux (CIUSSS) du Centre-sud de l'île de Montréal – Institut Raymond-Dewar, and the CIUSSS de la Capitale Nationale – Institut de réadaptation en déficience physique de Québec. Also, we want to thank the Faculté de médecine of the Université Laval, which offered a bursary to Audrey Clothier who completed a summer research internship under one of the researchers' supervision. The authors wish to thank audiologists Dale Bonnycastle and Ronald Choquette for the back-translations, translator Lorraine Paquet, and the five participants who contributed to the pre-test.

Authors' Note

Correspondence concerning this article should be addressed to Claude Vincent, Centre interdisciplinaire de recherche en réadaptation et intégration sociale, 525 boulevard Hamel, Office H-1112, Québec (QC), G1M 2S8. Canada. Email: **Claude.Vincent@rea.ulaval.ca.**

Appendix A. Hearing Loss Stigma (HLS) Questionnaire (Adapted from King, 2007)

NAME: _____ DATE: _____

DATE OF BIRTH: _____ MALE FEMALE

Instructions:

You will find below a list of sentences. For each one of them, you need to check off the answer that best suits you by circling the answer in the appropriate square.

Answer all the questions without exception. Don't spend too much time thinking about the answer, as it is your first impression that is important.

1. I have been discriminated against in education because of my hearing problems.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree

2. Sometimes I feel that I am being talked down to because of my hearing problems.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree

3. Having had hearing problems has made me a more understanding person.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree

4. I do not feel bad about having hearing problems.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree

5. I worry about telling people I received help concerning my hearing problems. *

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree

6. Some people with hearing problems are cognitively challenged. *

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree

7. People have been understanding of my hearing loss.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree

8. I have been discriminated against by friends and relatives because of my hearing problems.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree

9. I have been discriminated against by employers because of my hearing problems.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree

10. My hearing problems have made me more accepting of other people.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree

11. Very often I feel alone because of my hearing problems.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree

12. I am scared of how other people will react if they find out about my hearing problems.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree

13. I would have had a better chance in life if I did not have hearing problems.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree

14. I do not mind people in my neighborhood knowing I have hearing problems.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree

15. I would say I have hearing problems if I was applying for a job.*

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree

16. I worry about telling people that I have consulted hearing experts about my hearing problems.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree

17. People's reactions to my hearing problems make me keep myself to myself.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree

18. I am angry with the way people have reacted to my hearing problems.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree

19. I have not had any trouble from people because of my hearing problems.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree

20. I have been discriminated against by health professionals because of my hearing problems.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree

21. People have avoided me because of my hearing problems.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree

22. People have insulted me because of my hearing problems.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree

23. Having hearing problems has made me a stronger person.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree

24. I do not feel embarrassed because of my hearing problems.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree

25. I avoid telling people about my hearing problems.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree

26. Having hearing problems makes me feel that life is unfair.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree

27. I feel the need to hide my hearing problems from my friends.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree

28. I find it hard telling people I have hearing problems.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree

* The temporal reproducibility of these items (# 5, 6 and 15) was not observed. Do not take this into account during longitudinal follow-up.

Appendix B. Stigmatisation associée à la déficience auditive or HLS-CF[1]

NOM : _____ DATE : _____

DATE DE NAISSANCE : _____ HOMME FEMME

Directives :

Voici une liste de phrases. Pour chacune, encerclez la réponse qui correspond le mieux à votre expérience.

Veuillez indiquer une réponse pour chacune des phrases sans passer trop de temps à analyser les choix de réponse: c'est votre première impression qui importe.

1. J'ai subi la discrimination durant mes études en raison de mes problèmes d'audition.

Complètement en accord	En accord	Ni en accord ni en désaccord	En désaccord	Complètement en désaccord

2. J'ai parfois l'impression qu'on me rabaisse en raison de mes problèmes d'audition.

Complètement en accord	En accord	Ni en accord ni en désaccord	En désaccord	Complètement en désaccord

3. Mes problèmes d'audition ont fait de moi une personne plus compréhensive.

Complètement en accord	En accord	Ni en accord ni en désaccord	En désaccord	Complètement en désaccord

4. Je ne m'en fais pas à propos de mes problèmes d'audition.

Complètement en accord	En accord	Ni en accord ni en désaccord	En désaccord	Complètement en désaccord

5. Je crains de dire aux gens que j'ai reçu de l'aide pour mes problèmes d'audition. *

Complètement en accord	En accord	Ni en accord ni en désaccord	En désaccord	Complètement en désaccord

6. Certaines personnes ayant des problèmes d'audition ont des difficultés cognitives. *

Complètement en accord	En accord	Ni en accord ni en désaccord	En désaccord	Complètement en désaccord

7. Les gens se montrent compréhensifs à l'égard de mes problèmes d'audition.

Complètement en accord	En accord	Ni en accord ni en désaccord	En désaccord	Complètement en désaccord

[1]Hearing Loss Stigma (HLS) Questionnaire in Canadian French

8. Je subis la discrimination de la part de mes amis et parents en raison de mes problèmes d'audition.

Complètement en accord	En accord	Ni en accord ni en désaccord	En désaccord	Complètement en désaccord

9. Je subis la discrimination de la part d'employeurs en raison de mes problèmes d'audition.

Complètement en accord	En accord	Ni en accord ni en désaccord	En désaccord	Complètement en désaccord

10. Mes problèmes d'audition ont fait de moi une personne qui accepte mieux les autres.

Complètement en accord	En accord	Ni en accord ni en désaccord	En désaccord	Complètement en désaccord

11. Il m'arrive très souvent de me sentir seul(e) en raison de mes problèmes d'audition.

Complètement en accord	En accord	Ni en accord ni en désaccord	En désaccord	Complètement en désaccord

12. Je crains la façon dont les autres personnes réagiront si elles découvrent mes problèmes d'audition.

Complètement en accord	En accord	Ni en accord ni en désaccord	En désaccord	Complètement en désaccord

13. J'aurais eu plus de chance dans la vie si je n'avais pas eu de problèmes d'audition.

Complètement en accord	En accord	Ni en accord ni en désaccord	En désaccord	Complètement en désaccord

14. Cela ne me dérange pas que les gens de mon voisinage soient au courant de mes problèmes d'audition.

Complètement en accord	En accord	Ni en accord ni en désaccord	En désaccord	Complètement en désaccord

15. Je dévoilerais mes problèmes d'audition si je postulais un emploi. *

Complètement en accord	En accord	Ni en accord ni en désaccord	En désaccord	Complètement en désaccord

16. Je crains de dire aux gens que j'ai consultés des spécialistes pour mes problèmes d'audition.

Complètement en accord	En accord	Ni en accord ni en désaccord	En désaccord	Complètement en désaccord

17. Les réactions des gens à mes problèmes d'audition m'amènent à ne pas me livrer.

Complètement en accord	En accord	Ni en accord ni en désaccord	En désaccord	Complètement en désaccord

18. La façon dont les gens réagissent à mes problèmes d'audition me met en colère.

Complètement en accord	En accord	Ni en accord ni en désaccord	En désaccord	Complètement en désaccord

19. Les gens ne m'ont jamais causé d'ennuis en raison de mes problèmes d'audition.

Complètement en accord	En accord	Ni en accord ni en désaccord	En désaccord	Complètement en désaccord

20. Je subis la discrimination de la part de professionnels de la santé en raison de mes problèmes d'audition.

Complètement en accord	En accord	Ni en accord ni en désaccord	En désaccord	Complètement en désaccord

21. Des gens m'ont évité(e) en raison de mes problèmes d'audition.

Complètement en accord	En accord	Ni en accord ni en désaccord	En désaccord	Complètement en désaccord

22. Des gens m'ont insulté(e) en raison de mes problèmes d'audition.

Complètement en accord	En accord	Ni en accord ni en désaccord	En désaccord	Complètement en désaccord

23. Mes problèmes d'audition ont fait de moi une personne plus forte.

Complètement en accord	En accord	Ni en accord ni en désaccord	En désaccord	Complètement en désaccord

24. Je ne suis pas gêné(e) de mes problèmes d'audition.

Complètement en accord	En accord	Ni en accord ni en désaccord	En désaccord	Complètement en désaccord

25. J'évite de dévoiler mes problèmes d'audition aux gens.

Complètement en accord	En accord	Ni en accord ni en désaccord	En désaccord	Complètement en désaccord

26. Mes problèmes d'audition m'amènent à penser que la vie est injuste.

Complètement en accord	En accord	Ni en accord ni en désaccord	En désaccord	Complètement en désaccord

27. Je sens le besoin de cacher mes problèmes d'audition à mes amis.

Complètement en accord	En accord	Ni en accord ni en désaccord	En désaccord	Complètement en désaccord

28. J'ai de la difficulté à dévoiler mes problèmes d'audition aux gens.

Complètement en accord	En accord	Ni en accord ni en désaccord	En désaccord	Complètement en désaccord

* La reproductibilité temporelle de ces items (#5, 6 et 15) n'a pas été observée. Ne pas prendre en compte cet élément lors d'un suivi longitudinal.

Appendix C. Stigma Consciousness Questionnaire (SCQ) for Hearing Loss (Adapted from Pinel, 1999)

NAME: _____ DATE: _____

DATE OF BIRTH: _____ MALE FEMALE

Instructions

Please circle the letter that indicates the extent to which you agree with each statement listed below.

A. Completely agree
B. Agree
C. More or less agree
D. Neither agree nor disagree
E. More or less disagree
F. Disagree
G. Completely disagree

1. Stereotypes about hearing loss have not affected me personally.	A B C D E F G
2. I never worry that my behaviors will be viewed as stereotypical of a person who has a hearing loss.	A B C D E F G
3. When interacting with people who have normal hearing, I feel like they interpret all my behaviors in terms of the fact that I have a hearing loss.	A B C D E F G
4. Most people with normal hearing do not judge people with hearing loss on the basis of their ability to hear.	A B C D E F G
5. My being hearing impaired does not influence how people with normal hearing act with me.	A B C D E F G
6. I almost never think about the fact that I have a hearing loss when I interact with people who have normal hearing.	A B C D E F G
7. My being hearing impaired does not influence how people act with me.	A B C D E F G
8. Most people with normal hearing have a lot more prejudicial thoughts about people with hearing loss than they actually express.	A B C D E F G
9. I often think that people with normal hearing are unfairly accused of having prejudicial thoughts about people with hearing loss. *	A B C D E F G
10. Most people with normal hearing have a problem viewing people with hearing loss as equals.	A B C D E F G

* The temporal reproducibility of this item (# 9) was not observed. Do not take this into account during longitudinal follow-up.

Appendix D. Échelle de la conscience de la stigmatisation personnelle (associée à la perte d'audition) or SCQ-CF[1] for Hearing Loss

NOM : _____ DATE : _____

DATE DE NAISSANCE : _____ Homme Femme

Directives

Pour chacun des énoncés ci-dessous indiquez votre niveau d'accord ou de désaccord

A. Complètement en accord
B. En accord
C. Plus ou moins en accord
D. Ni en accord ni en désaccord
E. Plus ou moins en désaccord
F. En désaccord
G. Complètement en désaccord

1. Les stéréotypes concernant la perte d'audition ne m'affectent pas personnellement.	A B C D E F G
2. Je ne suis jamais inquiet que mes comportements puissent être perçus comme étant typiques d'une personne ayant une perte d'audition.	A B C D E F G
3. Lorsque je dialogue avec des gens ayant une audition normale, je sens qu'ils interprètent tous mes comportements en fonction du fait que j'ai une perte d'audition.	A B C D E F G
4. La plupart des gens ayant une audition normale ne jugent pas les gens ayant une perte d'audition en fonction de leur capacité à entendre.	A B C D E F G
5. Le fait que je suis une personne malentendante n'a aucune influence sur la façon dont les gens ayant une audition normale agissent avec moi.	A B C D E F G
6. Je ne pense presque jamais au fait que j'ai une perte d'audition lorsque je dialogue avec quelqu'un qui a une audition normale.	A B C D E F G
7. Le fait que je suis une personne malentendante n'a aucune influence sur la façon dont les gens agissent avec moi.	A B C D E F G
8. La plupart des gens ayant une audition normale ont beaucoup plus de préjugés à l'endroit des personnes ayant une perte d'audition qu'ils ne le disent en réalité.	A B C D E F G
9. Je crois souvent que les personnes ayant une audition normale sont injustement accusées d'avoir des préjugés envers les personnes qui ont une perte d'audition. *	A B C D E F G
10. La plupart des gens qui ont une audition normale ont de la difficulté à considérer les personnes qui ont une perte d'audition comme étant des égaux.	A B C D E F G

* La reproductibilité temporelle de cet item (#9) n'a pas été observée. Ne pas prendre en compte cet élément lors d'un suivi longitudinal.

[1]Stigma Consciousness Questionnaire (SCQ) for Hearing Loss in Canadian French

Students' Evaluation of Audiology Simulation Training

L'évaluation des étudiants d'une formation en audiologie utilisant des mises en situation

KEY WORDS

AUDIOLOGY
CASE SCENARIOS
DEBRIEFING
EDUCATION
EVALUATION
MANIKINS
SIMULATION
STANDARDIZED PATIENTS
STANDARDIZED PARENTS

Ahmad A. Alanazi
Nannette Nicholson

Ahmad A. Alanazi,
King Saud bin Abdulaziz University for Health Sciences, Riyadh, SAUDI ARABIA

Nannette Nicholson,
University of Arkansas for Medical Sciences, Little Rock, Arkansas, USA

Abstract

The use of simulation in the field of audiology as a strategy and tool for teaching and learning in clinical education programs is increasing. Eliciting feedback from students is important to design, improve, and implement successful simulation learning experiences. Yet, few simulation studies have reported outcomes of student feedback following simulation training. The purpose of this study was to explore students' perceptions of the simulation training components following 3 simulated hearing screening and parent counselling scenarios. Seventeen Doctor of Audiology (Au.D.) students participated in a simulation training, which included the use of a manikin, standardized parents, 3 case scenarios, debriefing sessions, and assessment. This cross-sectional mixed-methods study used a 12-item survey to elicit feedback from the students' perspective about simulation training components. This survey consisted of 10 statements with a Likert scale rating response methodology (1 = strongly disagree, 7 = strongly agree) and 2 open-ended questions to elicit written comments. Participants completed the feedback perception tool after the final case scenario. Overall, students agreed or strongly agreed (M = 6.74, SD = 0.32) that the simulation event enhanced their learning experience and opportunities for quality improvement were identified. Results showed student appreciation and recognition of the simulation training as adding value and enhancing their learning experience. Attention to details, organization, adequate time, participants' feedback, and evaluation when planning and preparing simulation training is one way to achieve higher participant satisfaction levels. Additional research on student perception of simulation training components will provide evidence to inform future simulation training.

Abrégé

Il y a une augmentation de l'utilisation de mises en situation comme stratégie et outil de formation et d'apprentissage dans les programmes d'enseignement clinique du domaine de l'audiologie. Une rétroaction de la part des étudiants est importante pour la conception, l'amélioration et la réussite de l'implantation d'expériences d'apprentissage utilisant des mises en situation. À ce jour, peu d'études ont recueilli la rétroaction d'étudiants ayant participé à des mises en situation. L'objectif de cette étude était d'explorer la perception des étudiants à l'égard de diverses composantes de mises en situation cliniques, et ce, après qu'ils aient participé à trois scénarios simulant des dépistages auditifs et du *counseling* à des parents. Dix-sept étudiants au doctorat en audiologie ont participé à une formation utilisant des mises en situation et comprenant l'usage d'un mannequin, de « patients simulés », de trois scénarios de cas, de périodes de discussion guidée entre le participant et l'animateur (*debriefing sessions*) et d'évaluations. Cette étude transversale à méthodes mixtes a utilisé un sondage composé de 12 items pour recueillir la rétroaction des étudiants concernant les différentes composantes d'une formation utilisant des mises en situation. Ce sondage comprenait 10 énoncés utilisant une échelle de Likert (1 = fortement en désaccord, 7 = fortement en accord) et deux questions ouvertes pour susciter des commentaires écrits. Les participants ont complété le sondage à la fin du troisième scénario. De façon générale, les étudiants ont indiqué qu'ils étaient en accord ou fortement en accord (M = 6.74, ET = 0.32) avec le fait que les mises en situation avaient optimisé leur expérience d'apprentissage et ils ont identifié des améliorations potentielles de qualité. Les résultats ont montré que les étudiants appréciaient et reconnaissaient la valeur ajoutée d'une formation utilisant des mises en situation sur leurs apprentissages. L'attention portée aux détails, à l'organisation, à la durée, à la rétroaction des participants et à l'évaluation lors de la planification et de la préparation d'une formation utilisant des mises en situation sont plusieurs façons d'obtenir un taux de satisfaction plus élevé de la part des participants. Des recherches supplémentaires recueillant la perception des étudiants à l'égard des composantes d'une formation utilisant des mises en situation fournira des évidences afin de façonner les futures formations utilisant cette méthode d'apprentissage.

Simulation is one of the most valuable innovations in clinical education, and is defined as "an act of imitating the behavior of a physical or abstract system, such as an event, situation or process that does or could exist" (Baek, 2009, p. 27). Simulation has become an accepted strategy in clinical education and training for healthcare professionals for two reasons: (a) increased attention to and emphasis on patient safety, and (b) evidence-based efficiency of simulation as a learning experience. Simulation training improves technical skills (Cook, 2014; Karakus, Duran, Yavuz, Altintop, & Caliskan, 2014; Ohtake, Marchilene, Schillo, & Rosen, 2013) and non-technical skills, such as critical thinking and decision-making (Lapkin & Levette-Jones, 2011; Wotton, Davis, Button, & Kelton, 2010). Simulation supports student practice with no fear of harming patients, thus reducing error and anxiety (Dearmon et al., 2013; Yule, Flin, Paterson-Brown, & Maran, 2006). The use of simulation as a learning environment is an innovative method for training audiology students; however, the use of simulation in audiology is still in its earliest stages (Alanazi et al., 2016). Simulation training can be divided into two categories: (a) simulation environment and (b) learning experience. These categories consist of several important components, such as manikins, safe environments, case scenarios, standardized patients (SPs), facilitators, debriefing, and students' satisfaction.

Simulation environment

The simulation environment is a physical place where simulation training is conducted and where the facilitator creates a friendly learning atmosphere (i.e., a safe environment), focuses on the learning objectives, and manages time (Fanning & Gaba, 2007; Meakim et al., 2013; Rall, Manser, & Howard, 2000). The simulation facility requires space, staff (e.g., facilitators and technicians), technology (e.g., video-playback systems and cameras), roles, objectives, time allocation, manikins with different fidelities (i.e., low, mid, or high fidelities), observing and debriefing rooms, adequate funding, access to SPs, etc. Orientation to the simulation environment before simulation training is also a critical part of creating the safe environment. All of these requirements help in providing successful educational experiences.

Although "simulation is a technique, not technology" (Gaba, 2007, p. 126), simulation training often depends on manikins. The use of manikins can enhance the students' learning experience because of their advanced capabilities and outputs, such as physiological changes (Epps, White, & Tofil, 2014). Manikins have been successfully used in both learning and assessment of clinical skills to achieve many learning objectives (Blackstock & Jull, 2007). The simulation accuracy of imitating reality determines the level (i.e., low, mid, or high) of manikin fidelity (Issenberg & Scalese, 2008; Wu & Shea, 2009). Low-fidelity manikins are frequently used in medicine because of their lower cost and the potential for repetitive use (Grober et al., 2004). A common misconception reported in the literature is that a high-fidelity simulation is better than a low-fidelity one. High-fidelity simulation is useful for skills involving interactions between students' cognitive and hands-on skills, as well as interaction with other healthcare personnel in the same simulation training (Gaba, 2006). Maran and Glavin (2003) suggest that manikins, regardless of their fidelity, are almost all potentially useful, but because of a lack of clear educational goals, many manikins are insufficiently used.

Manikins can be either controlled by an operator (e.g., a facilitator), or are automated (i.e., autonomous), changing status according to the intervention (Epps et al., 2014). The use of manikins as a teaching and assessment tool has recently been reported in the field of audiology (Alanazi et al., 2016; Kaf, Masterson, Dion, Berg, & Abdelhakiem, 2013). However, few manikins are available to train audiology students. For example, Baby Isao, manufactured by Intelligent Hearing Systems (2016), is a high-fidelity manikin that can be used to teach infant hearing screening and diagnostic techniques (i.e., otoacoustic emissions [OAEs] and auditory brainstem responses [ABRs]). OAEs are sounds emitted by the cochlea, either spontaneously or evoked by an auditory stimulus. ABRs are neuroelectrical signals (or auditory evoked potentials) generated by the auditory nerve and brainstem in response to an auditory stimulus. The simulator used in the current study consisted of the Baby Isao doll, the simulator box, a laptop computer, and software.

Learning experience

Standardized patients (SPs). SPs are trained actors who mimic or present particular scenarios. Prior to the use of SPs, training and evaluating healthcare students was performed by observing students' clinical skills with real patients (Stillman et al., 1986). This method was not efficient due to the differences between patients in terms of symptoms and other situational factors such as appointment time, attendance, and difficulties with accommodation of all students to observe one case. Thus, other training and evaluation methods were developed to assess healthcare students' skills, one of which is the use of SPs (Howley, 2013). The use of SPs has become one of the most common forms of physical examination and communication skills assessments in medical education (Epstein & Hundert, 2002). In audiology, there is shortage of

published studies that use SPs as an educational method (Alanazi et al., 2016).

Case scenario. The case scenario structure and content depend on the purpose of using SPs (i.e., the goal of the SP encounter). If the learning expectations of the simulation training are set at high learning levels (e.g., students implement all the core competencies of interprofessional collaborative practice: (a) values and ethics, (b) roles and responsibilities, (c) interprofessional communication, and (d) teams and teamwork), then a detailed case scenario is needed to meet all of the objectives (Howley, 2013; Interprofessional Education Collaborative Expert Panel, 2011). Although efforts to develop a guide for preparing SP case scenarios in healthcare simulation have been proposed (Baile et al., 2000; Cahill, 2015; Kim et al., 2006; Seropian, 2003), there remains a shortage of developed SP cases and related materials in the literature (Howley, 2013). For example, Seropian (2003) suggests that case scenarios include several elements: (a) objectives, (b) personnel and equipment, (c) computer setup and operator instructions, (d) paperwork and supporting documentation, (e) context, (f) knowledge and teaching information, (g) references related to the objectives, and (h) notes for further improvement of the scenarios. Kim et al. (2006) report that case scenarios should be: (a) relevant, (b) realistic, (c) engaging, (d) challenging, and (e) instructional. Generally, SPs could be involved in the simulation training in three ways: (a) the pre-encounter stage, where information about the SP is given to the student before the actual encounter; (b) the encounter stage, where the student meets the SP; and (c) the post-encounter stage, where feedback is given to the student by the SP (Dinsmore, Bohnert, & Preminger, 2013).

Debriefing. Debriefing is a process following the simulation exercise consisting of a guided discussion between facilitators and participants in an effort to enhance understanding of what went well and what could have gone better during the simulation exercise. Debriefing helps participants connect what they have learned in the simulation training with previous knowledge to enhance their learning (Fanning & Gaba, 2007). There is no standard structure of the debriefing process; nevertheless, several models have been proposed to help educators organize the structure, such as the Guidelines, Recommendations, Events, Analysis, and Transfer (GREAT) model and the Promoting Excellence and Reflective Learning in Simulation (PEARLS) framework (Dufrene & Young, 2013; Eppich & Cheng, 2015; Owen & Follows, 2006). The PEARLS framework specifies four distinct phases of the debriefing process, including: (a) reactions, (b) description, (c) analysis, and (d) summary. This approach focuses on identifying positive aspects of the training (what went well) as well as negative aspects (what could have gone better), while eliciting suggestions regarding aspects they would change if given another opportunity. The goal is to use context-specific factors, including choice of approach, time availability, students' rationale for action, and learning/performance gap between objectives and knowledge, skills, or behaviours, thereby facilitating and maximizing clinical decision-making (Eppich & Cheng, 2015).

Debriefing is still considered the underdeveloped part of simulation training (Neill & Wotton, 2011). Participation in debriefing is expected to increase the participants' ability to transfer knowledge to real situations (Halm, Lee, & Franke, 2011). For example, Ryoo and Ha (2015) explored the effect between the use and non-use of debriefing on clinical performance competency among 49 second-year nursing students. They found that the debriefing group (n = 24) scored significantly higher than the non-debriefing group (n = 25) in communication skills and in another 15 skills in the psychomotor domain. Similarly, Shinnick, Woo, Horwich, and Steadman (2011) examined the difference in knowledge of heart failure among 162 students who were assigned into debriefed and non-debriefed groups. Debriefed students showed an increase in knowledge of heart failure. Morgan et al. (2009) divided 71 anesthesiologists into two groups (debriefed and non-debriefed) and found that the non-debriefed group scored lower on technical skills.

Facilitator. The role of the facilitator can be filled by a trained simulation facilitator, faculty member, or student, depending on the level of facilitation needed: high, intermediate, or low (Fanning & Gaba, 2007). The debriefing process and role of the facilitator are integrally related. While the literature suggests using debriefing as an integrated component of healthcare simulation training, few studies report outcomes of the debriefing process or debriefing practices, particularly in audiology (Alanazi, Nicholson, & Thomas, 2017). Fanning and Gaba (2007) stated, "There are surprisingly few papers in the peer-reviewed literature to illustrate how to debrief, how to teach or learn to debrief, what methods of debriefing exist, and how effective they are at achieving learning objectives and goals" (p. 115). Recognizing this gap in information, Lusk and Fater (2013) explored the debriefing process and role of the facilitator and debriefing process across disciplines such as aviation, psychology, education, medicine, and nursing, and identified common themes and practices. A common practice is the use of Tanner's model of clinical judgment to facilitate critical thinking and clinical decision-

making skills. This model incorporates four phases, including: (a) noticing, (b) interpreting, (c) responding, and (d) reflecting. This continuous cycle of moving in and out of phases (reflection-in-action and reflection-on-action) provides students with opportunities to practice generalization and application of clinical judgment (Lusk & Fater, 2013). Thus, the debriefing process within the simulation training session serves as a platform to coach and assist students as they learn to apply and generalize skills. The PEARLS framework of facilitation can be used in conjunction with Tanner's model of clinical judgment to optimize student learning outcomes and skill development.

Why students' perception of the simulation training is important

Training students in the simulation facilities needs to be meaningful for students. The use of evaluation tools of students' perceptions is a method to increase meaning, deepen the learning experience, gather more information about student preferences, and plan for quality improvement of the simulation training. Implementation of evidence-based educational practices requires an approach in which current, high-quality, rigorous research evidence is integrated with educator expertise and student preferences (Coalition for Evidence-Based Policy, 2003). Therefore, the evaluation of the simulation training by students is critical in building and designing successful simulation training (or simulation programs).

The evaluation of the simulated training differs from the assessment of students' performance and learning outcomes, which use assessment tools such as the Audiologic Counseling Evaluation (Adamson, Kardong-Edgren, & Willhaus, 2013; English, Naeve-Velguth, Rall, Uyehara-Isono, & Pittman, 2007). Many evaluation tools have been developed that focus on student self-reports of their perception and/or satisfaction with the simulation training (Alanazi et al., 2016; Alinier et al., 2008; Levett-Jones et al., 2011). While verbal debriefing is the more common procedure to facilitate learning following simulation training, Lestander, Lehto, and Engström (2016) suggest that the post-simulation evaluation serve as another opportunity for student reflection. Petranek (2000) suggests a written reflection as an efficient learning strategy, while Baikie and Wilhelm (2005) propose that written words facilitate expression of experiences that are too sensitive to describe face-to-face. The use of open-ended questions is recommended to generate new information that may have otherwise been overlooked (Knudsen et al., 2012). However, reports on the use of these assessment tools and/or the contribution of the results to quality improvement efforts in planning subsequent simulation training are limited.

Alanazi et al. (2017) conducted a systematic review of publications in health professions to identify and evaluate the best available evidence (level and quality) of the use of simulation training to improve clinical skills, knowledge, and self-confidence among healthcare students. The authors reported that only seven of 30 reviewed studies reported students' satisfaction. When all the simulation-training components are put together appropriately, a high level of satisfaction among participants is expected. Student participants' satisfaction is important in clinical education because it may correlate with performance and may help students develop skills and acquire knowledge (Bremner, Aduddell, Bennett, & VanGeest, 2006; Pike, 1991). Thus, the purpose of this study was to explore students' perceptions and satisfaction with the hearing screening and parent counselling simulation training.

Methods

This study was conducted at the University of Arkansas for Medical Sciences (UAMS) Simulation Center and received the UAMS Institutional Review Board approval (#204279). The simulation training consisted of pre-event exposure to knowledge, three case scenarios with specific objectives relevant to newborn hearing screening and parent counselling, and the combined use of Baby Isao with SPs in the role of standardized parents, who are in the position of making informed decisions that will impact their child's future (e.g., parents choose spoken or signed language as a method of communication for their child). The content and format of the simulation case scenarios used in this study are shown in Table 1 and have been previously described in detail by Alanazi and colleagues (2016).

Participants

Seventeen female Doctor of Audiology (Au.D.) students (*M* age = 24.59 years, *SD* = 1.50, range = 22–29 years; Au.D. cohort = second- and third-year students) participated as volunteers in this study. The role of students in the simulation training was either as active or passive (observer) participants. Six students (two students in active roles, one from each year in the program, per scenario) conducted the hearing screening and counselled the parents, whereas the remaining students (passive role) watched the case scenarios unfold on a large screen monitor through the closed-circuit video system. All students participated in the briefing and debriefing sessions.

Table 1. Standardized Parents and Case Scenarios

Standardized Parent	Case Scenario
One ethnically diverse standardized parent	The baby failed the screening. The mother accepted the results and refused the follow-up diagnostic evaluation due to religious and cultural beliefs.
Two standardized parents (culturally deaf in real life and in the scenario)	The baby passed the hearing screening. A certified sign language interpreter was recruited. The parents were unhappy because their baby passed the screening.
Two standardized parents	The baby failed the screening. The father was angry and blamed the mother, who was a musician, because she exposed the child to loud music in utero.

Simulation training

Two types of simulation were used in this study: (a) one manikin, Baby Isao, and (b) five trained standardized parents, and one sign-language interpreter representing three different case scenarios (Table 1). Each scenario consisted of a 10-minute briefing session, a 20-minute simulation experience (i.e., hearing screening and counselling parents), and a 30-minute debriefing session guided by an experienced facilitator.

Materials

A 12-item perception survey (Appendix A: Students' Perception of Simulation Training Components [SPSTC] survey) consisting of 10 statements and two open-ended questions was developed by the UAMS Simulation Center personnel to include the critical components of simulation training as discussed in the literature. This survey was modified by the authors to collect students' perceptions and feedback about this training through three aspects: (a) the simulation environment, (b) the learning experience, and (c) the highlights of the simulation training. Students were asked to rate their level of agreement with the simulation training categories based on a Likert scale, where 1 = strongly disagree (very dissatisfied) and 7 = strongly agree (very satisfied). Participants were instructed to use "not applicable" if a statement did not pertain to the simulation training performed. Statements rated as "not applicable" were not assigned a numeric value and were eliminated from the average ratings. Each participant was given an opportunity to provide short answers to inquiries about the third category (i.e., the highlights of the simulation training). The two open-ended inquiries designed to elicit additional information were: (a) "Describe any part of the simulation training that was exceptional" and (b) "Describe any part of the simulation training that did not meet your expectations".

Procedures

Student participants completed a pre-training curriculum about newborn hearing screening training on the National Center for Hearing Assessment and Management (2015) website and had observed 10 hours of neonatal hearing screening as part of their clinical rotations at Arkansas Children's Hospital prior to the simulation. In addition, students were given the opportunity to practice conducting hearing screening with Baby Isao on their own before the simulation event. Details about the upcoming simulation training were not provided prior to the event. On the day of the simulation event, two student volunteers were randomly selected by the facilitator prior to each case to perform the hearing screening and break bad news (e.g., a baby has a hearing loss) and counsel the standardized parents about the next steps in the process. The remaining student watched the simulated scenarios via a widescreen video monitor in a separate room. The debriefing sessions were structured using the PEARLS framework and were guided by a trained simulation facilitator familiar with the learning objectives. Tanner's model of clinical judgment was used to facilitate critical thinking and clinical decision-making skills (Lusk & Fater, 2013). Audiology faculty members participated in the briefing and debriefing sessions, and the standardized parents participated in the debriefing session in which they performed. The three case scenarios were completed sequentially in one day. The total simulation training was completed in about three hours. After the final case scenario, the SPSTC survey was distributed and students were asked to complete the evaluation of the simulation training prior to leaving the centre.

Results

Quantitative and qualitative results of the responses to the SPSTC survey are presented for three aspects of training: (a) simulation environment, (b) learning experience, and (c) highlights of the learning experience. Descriptive statistics are presented for items 1–10 and a thematic analysis is presented for items 11 and 12. Items 1–10 were rated using a Likert scale, where 1 = strongly disagree and 7 = strongly agree. Responses rated as not applicable were eliminated from the analysis. Overall, these results suggest that the majority of students agreed or strongly agreed that the simulation training event enhanced their learning experience (*M* = 6.74, *SD* = 0.32, range = 6–7). An additional analysis was conducted to explore specific feedback responses about the simulation environment (items 1–3), the learning experience (items 4–10), and highlights of the learning experience from the students' perspective (items 11–12).

Simulation environment

The overall mean for items 1–3 was 6.51 (range = 1–7, *SD* = 0.77), based on a Likert scale where 1 = strongly disagree and 7 = strongly agree. Figure 1 shows the mean student ratings for items 1 (suitability), 2 (well-equipped), and 3 (safety). Of the total students, two passive students (#1 and #5) strongly disagreed with the following statement: "The orientation to the simulation was suitable" (item 1). One student (#13) rated the statement "The simulation center was well equipped" (item 2) as not applicable. This student's response was eliminated from the analysis.

Learning experience

The overall mean of responses for this category (items 4–10) was 6.93 (range = 1–7, *SD* = 0.11). Student perceptions of the following items were elicited: item 4 = case scenario, 5 = debriefing, 6 = reflection, 7 = facilitator, 8 = standardized parents, 9 = feedback, and 10 = application. The statement "The learning experience will help me in my clinical practice" (item 10) was the only item rated with "strongly agree" by all students. Five passive student participants (#8, #10, #12, #13, and #16) rated "The debriefing sessions helped me reflect on my practice" statement (item 6) as not applicable.

Highlights of the simulation training

Although the amount of qualitative data (i.e., responses

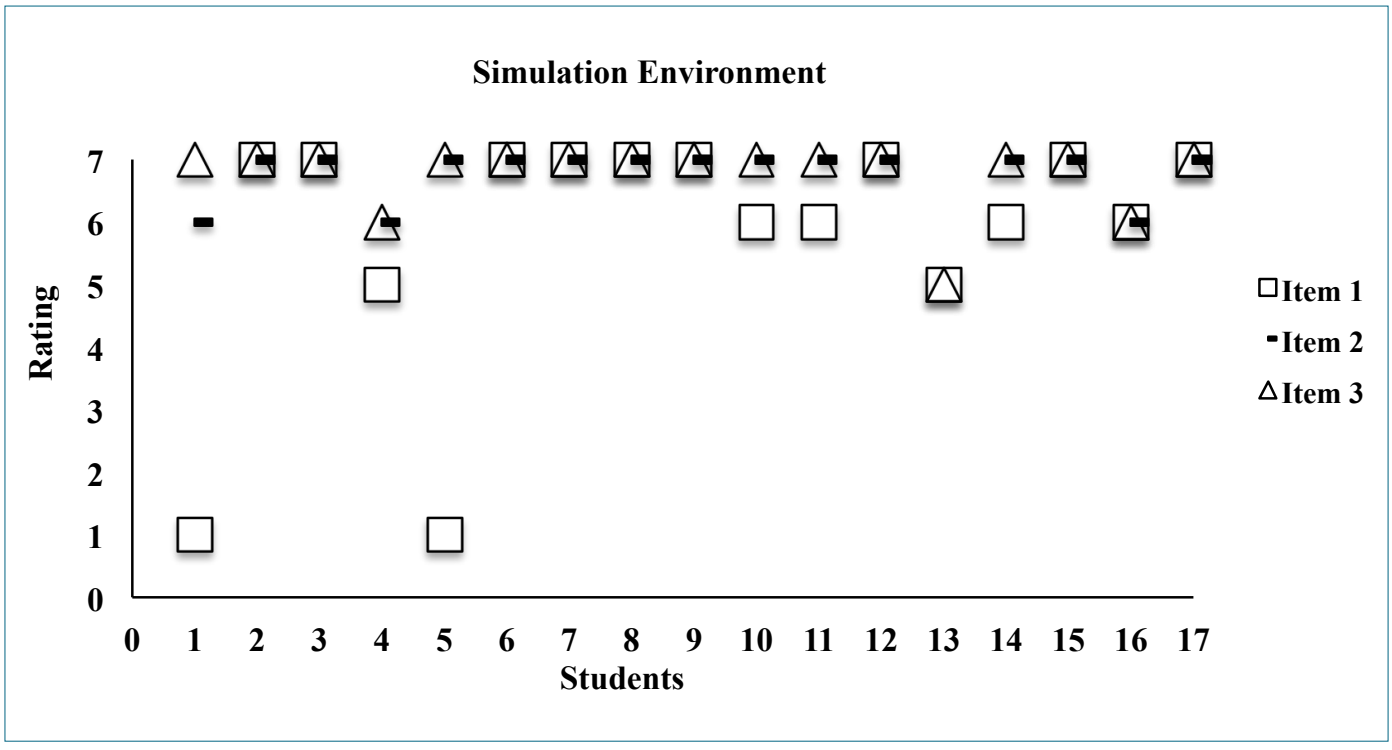

Figure 1. Simulation environment ratings for items 1–3 on the perception survey are shown by participant (*N* = 17). Simulation training ratings were 1 = strongly disagree to 7 = strongly agree. *Note*. Student #13 rated item 2 as not applicable, so there are only two responses. Item 1 = "The orientation to simulation was suitable"; item 2 = "The simulation center was well equipped"; item 3 = "The simulation environment felt safe for participation".

to open-ended statements: items 11 and 12) was not huge, these responses were imported into NVivo qualitative data analysis software (QSR International Pty Ltd., 2015). The frequency of thematic concepts was identified. The deductive qualitative content analysis (i.e., themes [simulation components] in this approach are already known from the survey) was used to explore these responses generated from the open-ended statements (Burnard, Gill, Stewart, Treasure, & Chadwick, 2008). An open coding procedure was performed by reading each response to these statements and making notes next to key words of the responses. The five most frequently presented themes are shown in Figure 2. The brackets within quotations are used to clarify meaning and provide a brief explanation.

More than half of student participants (53%) reported that all of the training components were exceptional, without referring to an individual component. For example, student #3 said, "Everything was wonderful. I really did not realize how valuable of an experience this would be." Student #7 said, "I loved this experience. It would be great to have the opportunity to have a rotation here for all of us in the future. If not this, more events like this would be great!" The remaining students reported individual components as the highlights of the simulation training. Six percent of the students indicated safety of the environment as the most exceptional component of the simulation training. For example, student #1 described the exceptional component of the simulation training as "Practice with counseling without affecting real patients." Of the total students, 17% identified the use of the standardized parents, 12% the case scenarios, and 12% the debriefing component as the highlights of the simulation training. For instance, student #11 stated, "The actors were exceptional." Student #2 reported, "The case scenarios were so realistic." Moreover, student #8 described the case scenarios as "Such realistic scenarios- ones that we do not see often and could use some hands on with!" Student #6 commented on debriefing with, "The debriefing helped a lot. I learned so much to take into my daily practice." Six percent of the students suggested that more structured briefing during the orientation session would be helpful to be familiarized with the simulation environment. One student noted that briefing did not meet her expectations: "We [students] need to know more before setting, what is expected of us [to do in the simulation environment]?"

Discussion

To evaluate the simulation training from the students' perspectives, a post-event evaluation survey was used to elicit feedback about three major components of the simulation training: (a) simulation environment, (b) learning experience, and (c) highlights of the simulation training. Results and findings from the current study suggest that the simulation training enhanced students' perception of the learning experience. These results are consistent with previous studies that reported students' satisfaction (Alanazi et al., 2016; Dearmon et al., 2013; Ohtake et al., 2013).

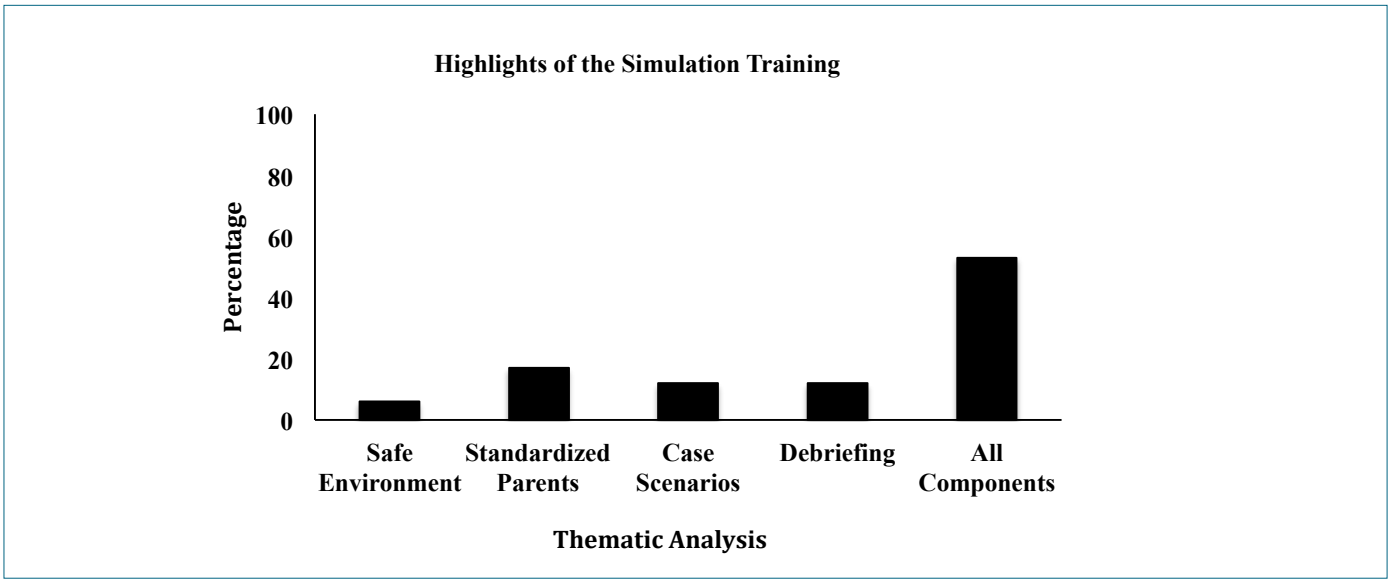

Figure 2. Simulation training components that were exceptional according to the students' descriptions and the number of participants who referred to each of the themes ($N = 17$). *Note.* "All Components" indicates the responses demonstrating that everything was exceptional about the simulation environment and learning experience.

Simulation environment

Orientation. The vast majority of simulation activities take place in simulation centres, teaching hospitals, and medical schools (Passiment, Sacks, & Huang, 2011). Orientation to the simulation environment before a simulation activity is necessary, because it allows students to become familiar with simulators, roles, objectives, and time allocation (Meakim et al., 2013). Lack of introduction may contribute to a feeling of anxiety and may leave students feeling underprepared and unable to apply the knowledge and practice the skills. In the current study, two students strongly disagreed with the statement, "The orientation to the simulation was suitable" (item 1). Although all students had completed the newborn hearing screening training module before the actual simulation training, no additional details about the upcoming event were given to students. Instructions were limited in that students knew they would be participating in an event at the UAMS Simulation Center with no further details about the event. Giving learners detailed information in advance about the patient's condition and what was going to happen in the encounter stage would (a) reduce the benefits of the simulation training because learners would lose the ability to understand the learning objectives by themselves, and (b) remove the element of surprise (Alinier, 2011). However, it is recommended that facilitators provide general learning objectives from which learners cannot predict exactly what will happen in the scenarios. Students' feedback presents faculty and facilitators with an opportunity for quality improvement in execution of the case scenarios in subsequent simulation training. In addition, lessons learned point toward the need to offer better general descriptions of the tasks that the students are expected to complete during future training sessions.

Equipment. Sixteen student participants in our study agreed or strongly agreed with the statement "The simulation center was well equipped" (item 2). One reason for the high rating of this component may be the fact that the current study was conducted in the UAMS Simulation Center. The Simulation Center contains seven simulation theatres fully equipped with high-fidelity manikins, overhead viewing cameras, panoramic wall-mounted units, and five debriefing classrooms. The use of Baby Isao, involving various patient states and background noise conditions for demonstration and simulation purposes to teach hearing screening, was unique. One student rated this item (item 2) as not applicable. The authors explored why this student might not have understood the relevance of this item, inappropriately marking it as not applicable. Since audiology is an equipment-intensive field, and this was the audiology student's first time participating in simulation training at the Simulation Center, it may be that the student misunderstood the statement and thought it referred to audiology equipment as opposed to the Simulation Center equipment or facility. One quality improvement modification may be to restate this item as "The simulation center facility was well-equipped". This rewording may help clarify the intent of this item.

Safety. The simulation training is a learning environment and should be physically comfortable (i.e., feeling safe and relaxed expressing oneself and emphasizing trust). Meakim et al. (2013) defined the safe training environment as "the emotional climate that facilitators create by the interaction between facilitators and participants. In this positive emotional climate, participants feel at ease taking risks, making mistakes, or extending themselves beyond their comfort zone" (p. S9). Without such an environment, the simulation training may be restricted to achieve its goals. All students in the current study rated the statement "The simulation environment felt safe for participation" (item 3) with "agree" or "strongly agree", indicating that they felt that it was a safe setting to practice and learn through action and interaction with the standardized parents.

Learning experience

Standardized patients. SPs are not intended to replace experience with real patients, but they are used to teach and evaluate clinical skills and knowledge in a safe environment (Barrows, 1993; Stroud, Smith, Edlund, & Erkel, 1999). Because of the numerous advantages of the use of SPs, many health professions have used SPs as a standard teaching approach; therefore, audiology programs are encouraged to use SPs as standard practice for their students. In the present study, the standardized parents, who were professional actors with prior paid experience, were included to train students on how to deliver bad news and counsel parents. Therefore, the standardized parents were reliable in imitating the case scenarios and provided participants with helpful advice. Patient feedback is important in terms of pointing out strengths and weaknesses of students' skills, and SPs offer this feedback from the patients' perspective (Howley & Martindale, 2004). This feedback is typically not available with real patients. Therefore, the standardized parents in our study participated in the debriefing session in which they performed. One of the standardized parents commented on active student participants, "There were a lot of points where you all definitely did things that put us at ease. Your tones of voice were very calming. And you all made really good eye contact." All student participants in

our study rated statements relevant to the standardized parents' performance (items 8 and 9) with "agree" or "strongly agree".

It is generally recognized that the use of SPs limits the number of active student participants that can be efficiently accommodated at a time (Bearnson & Wiker, 2005). In this study, the role of six students was active (i.e., they performed the hearing screening and encountered the standardized parents), while the role of the remaining students was passive (i.e., observation of the encounters with the standardized parents). Active students may have recognized more areas for improvement through active participation in the simulation sessions as opposed to passive participation. However, regardless of the role in the simulation training, all student participants strongly agreed that the learning experience was beneficial and would help them in their clinical practice (item 10). Comments provided in response to the open-ended questions indicated that some students believed that assignment to the active role would be a beneficial learning experience for all students. Quality improvement efforts will focus on implementation of this suggestion in future training.

Case scenarios. The case scenarios must reflect reality as much as possible. In the current study, all three stages of the use of the standardized parents in the simulation training (i.e., the pre-encounter, encounter, and post-encounter stages) were implemented. Moreover, three scenarios were designed to represent diverse cultural and socioeconomic backgrounds and incorporate a variety of emotional responses: an angry parent, parents from deaf culture experiencing grief, and a parent from a minority population displaying acceptance of hearing loss for cultural and religious beliefs. These scenarios required clinical judgment "in action" to quickly make a decision about the best way to respond to the situation. Following completion of the case, students were given an opportunity to use reflection "on action" about their choices and to discuss what went well, what did not go well, and what could have gone better. All students agreed or strongly agreed that the case scenarios seemed realistic (item 4). The detailed preparation, practice, and implementation of scripts contributed to the high satisfaction levels.

Debriefing. The structure of debriefing sessions is very important and can be achieved by using any of the debriefing models. Accrediting organizations such as the Council for Academic Accreditation (CAA), an organization under the American Speech-Language-Hearing Association (ASHA), require Au.D. programs to provide evidence that their students are able to demonstrate knowledge and skill (and professional competencies generally) in specific content areas (Council for Academic Accreditation, American Speech-Language-Hearing Association, 2016). Although ASHA may not consider debriefing hours as direct patient contact hours, this activity is equivalent to a "case conference" or review of a case following a clinical encounter, and can be recorded as hours for the "other" category. Decisions about how to count the time invested in debriefing activities following simulation are left to the interpretation and discretion of each accredited program.

In our study, the facilitator used the PEARLS model that helped to understand how and where students (a) expressed their feelings and thoughts, (b) described the learning experience, (c) followed a guided reflection, and (d) reviewed all the objectives through the facilitator guidance. The analysis phase of this model included a plus-delta analysis (+/Δ), in which the participants, observers, and the standardized parents reflected on the performance, including positive aspects (the +) as well as aspects they would change in the future (the Δ). Student participants in our study either agreed or strongly agreed with the statement "The debriefing sessions were well prepared" (item 5). Debriefing was rated by student participants as an exceptional component of their simulation training (item 5). Although simulation training research always refers to debriefing, attention to the systematic analysis of debriefing data is rare (Neill & Wotton, 2011; Wotton et al., 2010).

Furthermore, the reflection component has been used effectively as part of a pedagogical approach in audiology and communication sciences and disorders, and benefited students (Chabon & Lee-Wilkerson, 2006; Goldberg, Richburg, & Wood, 2006; Munoz & Jeris, 2005; Ng, Bartlett, & Lucy, 2012). The majority of student participants in our study rated the statement "The debriefing sessions helped me reflect on my practice" (item 6) with "agree" or "strongly agree". However, five students rated the same item as not applicable. Authors examined the student's roles as active or passive participants and found that the students rating the item as not applicable were passive participants.

Facilitator. An experienced debriefing facilitator may apply different techniques to guide the conversation and provide beneficial feedback. Moreover, the facilitator may create a friendly learning atmosphere, focus on the learning objectives, and manage time (Fanning & Gaba, 2007; Lederman, 1992; Rall et al., 2000). The perception of the simulation training is connected to the facilitator's skills (Fanning & Gaba, 2007). In high debriefing, the facilitator assists only if needed and the participants debrief themselves; in contrast, participants depend totally on the

facilitator in low debriefing. Intermediate debriefing requires less facilitator involvement than low debriefing. Our study required a high facilitation level because of the challenging scenarios incorporated in our simulation event. Student participants in our study either agreed or strongly agreed with the statement "The facilitator was supportive" (item 7).

Simulation training highlights

Repetitive practice is recognized as one of the best methods to facilitate learning (Bradley, 2006; Morey et al., 2003). Therefore, assessment of student satisfaction in simulation is important in terms of guiding quality improvement efforts for future training. Satisfaction does not equal increased knowledge and skill; however, correlation of students' perceptions and performance suggests that simulation may build self-confidence, which in turn helps students develop skills and acquire knowledge (Bremner et al., 2006). Alanazi and colleagues (2016) assessed Au.D. students' satisfaction after simulation training on hearing screening and parental counselling via a 23-item satisfaction survey. The authors reported that all participants rated their satisfaction level as "satisfied" or "very satisfied" after the educational simulation activity.

Open-ended statements were used in this survey as a mixed-method strategy to elicit qualitative student perception data about the exceptional features of this simulation training that may have been overlooked. This method provides immediate feedback and elicits relatively short immediate responses due primarily to the time constraints. However, this method fails to elicit the rich, thoughtful responses that are acquired without time constraints. Written reflections serve to facilitate critical thinking by providing students with the opportunity to connect previous experience with future actions based on lessons learned in the present (Petranek, 2000). Use of Tanner's model in a structured written reflection assignment could further enhance student learning by providing students with the opportunity to record their observations, interpret actions and decisions, analyze responses, and reflect upon outcomes and alternative scenarios (Lusk & Fater, 2013).

Addition of a written reflection assignment will be considered as a vital component in future simulation training to provide students with an opportunity to generalize and apply clinical judgment while at the same time providing a rich source of supplemental data from which to draw upon for quality improvement initiatives. Ng and colleagues (2012) provide an excellent tutorial on conducting qualitative research in audiology. Coupled with the use of a written reflection structured according to Tanner's model to assess clinical decision-making development and judgment in simulation, a qualitative study has the potential to generate new knowledge that may have been unnoticed with the time-constrained, open-ended feedback approach used in this study.

Limitations and future research

One limitation of this study is the small sample size. The convenience sample was composed of students enrolled in one Au.D. program. No attempt was made to control for participant ethnicity, gender, or age. The evaluation process was limited to the students' perspectives on the simulation training components and did not measure professional competencies in any specific area. Additional comparative information could have been gained by including faculty, facilitators, and the standardized parents in the assessment. Finally, the topic of this simulation training was broad enough that it could have easily been expanded in an academic health centre as an interprofessional training opportunity to include speech-language pathology, nursing, and other health profession students.

Conclusion

Assessment of the simulation training components in this study indicates that students viewed the use of the standardized parents, case scenarios, and debriefing using the PEARLS framework as a novel and effective approach in audiology education. Debriefing allows audiology students to reflect on their performance and feelings including positive aspects as well as aspects they would change in the future. In addition, evaluation was instrumental in identifying quality improvement opportunities for future simulation training, thus contributing to satisfaction with this training. The evaluation of participants' perceptions about simulation training is one way to achieve higher satisfaction (or agreement) levels when the same simulation training is repeated. Additional research on students' perceptions of simulation training components will provide evidence to inform future simulation training efforts, as well as facilitate the development and refinement of the perception survey used in this study.

References

Adamson, K. A., Kardong-Edgren, S., & Willhaus, J. (2013). An updated review of published simulation evaluation instruments. *Clinical Simulation in Nursing, 9*, e393–e400. doi: 10.1016/j.ecns.2012.09.004

Alanazi, A. A., Nicholson, N., Atcherson, S. R., Franklin, C., Anders, M., Nagaraj, N., ... Highley, P. (2016). Use of Baby Isao simulator and standardized parents in hearing screening and parent counseling education. *American Journal of Audiology, 25*, 211–223. doi: 10.1044/2016_AJA-16-0029

Alanazi, A. A., Nicholson, N., & Thomas, T. (2017). Use of simulation training to improve knowledge, skills, and confidence among healthcare students: A

systematic review. *The Internet Journal of Allied Health Sciences and Practice, 15*(3), 1–24.

Alinier, G. C. (2011). Developing high-fidelity health care simulation scenarios: A guide for educators and professionals. *Simulation & Gaming, 42*(1), 9–26. doi: 10.1177/1046878109355683

Alinier, G. C., Harwood, C., Harwood, P., Montague, S., Huish, E., & Ruparelia, K. (2008). *Development of a program to facilitate interprofessional simulation-based training for final year undergraduate health care students.* York, UK: Higher Education Academy. Retrieved from http://uhra.herts.ac.uk/handle/2299/4573

Baek, Y. (2009). Digital simulation in teaching and learning. In D. Gibson & Y. Baek (Eds.), *Digital simulations for improving education: learning through artificial teaching environments* (pp. 25–51). Hershey, PA: Information Science Reference.

Baikie, K. A., & Wilhelm, K. (2005). Emotional and physical health benefits of expressive writing. *Advances in Psychiatric Treatment, 11*(5), 338–346. doi: 10.1192/apt.11.5.338

Baile, W. F., Buckman, R., Lenzi, R., Glober, G., Beale, E. A., & Kudelka, A. P. (2000). SPIKES—A six-step protocol for delivering bad news: Application to the patient with cancer. *The Oncologist, 5*, 302–311. doi: 10.1634/theoncologist.5-4-302

Barrows, H. S. (1993). An overview of the uses of standardized patients for teaching and evaluating clinical skills. *Academic Medicine, 68*, 443–451.

Bearnson, C. S., & Wiker, K. M. (2005). Human patient simulators: A new face in baccalaureate nursing education at Brigham Young University. *Journal of Nursing Education, 44*, 421–425.

Blackstock, F. C., & Jull, G. A. (2007). High-fidelity patient simulation in physiotherapy education. *Australian Journal of Physiotherapy, 53*(1), 3–5. doi: 10.1016/S0004-9514(07)70056-9

Bradley, P. (2006). The history of simulation in medical education and possible future directions. *Medical Education, 40*, 254–262.

Bremner, M., Aduddell, K., Bennett, F., & VanGeest, J. (2006). The use of human patient simulators: Best practices with novice nursing students. *Nurse Educator, 31*(4), 170–174.

Burnard, P., Gill, P., Stewart, K., Treasure, E., & Chadwick, B. (2008). Analyzing and presenting qualitative data. *British Dental Journal, 204*(8), 429–432. doi: 10.1038/sj.bdj.2008.292

Cahill, M. S. (2015). Perspectives on the use of standardized parents to teach collaboration to graduate occupational therapy students. *American Journal of Occupational Therapy, 69*(2), 1–7. doi: 10.5014/ajot.2015.017103

Coalition for Evidence-Based Policy. (2003). *Identifying and implementing educational practices supported by rigorous evidence: A user-friendly guide.* Retrieved from https://www2.ed.gov/rschstat/research/pubs/rigorousevid/rigorousevid.pdf

Cook, D. A. (2014). How much evidence does it take? A cumulative meta-analysis of outcomes of simulation-based education. *Medical Education, 48*, 750–760. doi: 10.1111/medu.12473

Council for Academic Accreditation, American Speech-Language-Hearing Association. (2016). *Standards for accreditation of graduate education programs in audiology and speech-language pathology* (pp. 1–42). Retrieved from http://caa.asha.org/wp-content/uploads/Accreditation-Standards-2017.pdf

Chabon, S. S., & Lee-Wilkerson, D. (2006). Use of journal writing in the assessment of CSD students' learning about diversity: A method worthy of reflection. *Communication Disorders Quarterly, 27*(3), 146–158. doi: 10.1177/15257401060270030301

Dearmon, V., Graves, R. J., Hayden, S., Mulekar, M. S., Lawrence, S. M., Jones, L., … Farmer J. E. (2013). Effectiveness of simulation-based orientation of baccalaureate nursing students preparing for their first clinical experience. *Journal of Nursing Education, 52*(1), 29–38. doi: 10.3928/01484834-20121212-02

Dinsmore, B. F., Bohnert, C. & Preminger, J. E. (2013). Standardized patients in audiology: A proposal for a new method of evaluating clinical competence. *Journal of American Academy of Audiology, 24*, 372–392. doi: 10.3766/jaaa.24.5.5

Dufrene, C., & Young, A. (2013). Successful debriefing—Best methods to achieve positive learning outcomes: A literature review. *Nurse Education Today, 34*(3), 372–376. doi: 10.1016/j.nedt.2013.06.026

English, K. M., Naeve-Velguth, S., Rall, E., Uyehara-Isono, J., & Pittman, A. (2007). Development of an instrument to evaluate audiologic counseling skills. *Journal of the American Academy of Audiology, 18*(8), 675–687. doi: 10.3766/jaaa.18.8.5

Eppich, W., & Cheng, A. (2015). Promoting excellence and reflective learning in simulation (PEARLS): Development and rationale for a blended approach to health care simulation debriefing. *Simulation in Healthcare, 10*(2), 106–115. doi: 10.1097/SIH.0000000000000072

Epps, C., White, M. L., & Tofil, N. (2014). Mannequin based simulators. In A. I. Levine, S. DeMaria, A. D. Schwartz, & A. J. Sim (Eds.), *The comprehensive textbook of healthcare simulation* (pp. 209–232). New York, NY: Springer.

Epstein, R. M., & Hundert, E. M. (2002). Defining and assessing professional competence. *Journal of the American Medical Association, 287*(2), 226–235. doi: 10.1001/jama.287.2.226

Fanning, R. M., & Gaba, D. M. (2007). The role of debriefing in simulation-based learning. *Simulation in Healthcare, 2*(2), 115–125. doi: 10.1097/SIH.0b013e3180315539

Gaba, D. (2007). The future vision of simulation in healthcare. *Simulation in Healthcare, 2*(2), 126–135. doi: 10.1136/qhc.13.suppl_1.i2

Gaba, D. (2006). What does simulation add to teamwork training? *Agency for Healthcare Research and Quality*. Retrieved from https://psnet.ahrq.gov/perspectives/perspective/20

Goldberg, L. R., Richburg, M. C., & Wood, L. A. (2006). Active learning through service-learning. *Communication Disorders Quarterly, 27*(3), 131–145. doi: 10.1177/15257401060270030201

Grober, E. D., Hamstra, S. J., Wanzel, K. R., Reznick, R. K., Matsumoto, E. D., Sidhu, R. S., & Jarvi, K. A. (2004). The educational impact of bench model fidelity on the acquisition of technical skill: The use of clinically relevant outcome measures. *Annals of Surgery, 240*(2), 374–381. doi: 10.1097/01.sla.0000133346.07434.30

Halm, B. M., Lee, M. T., & Franke, A. A. (2011). Improving toxicology knowledge in preclinical medical students using high-fidelity patient simulators. *Hawaii Medical Journal, 7*, 122–115.

Howley, L. D. (2013). Standardized patients. In A. I. Levine, S. DeMaria, A. D. Schwartz, & A. J. Sim (Eds.), *The comprehensive textbook of healthcare simulation* (pp. 173–190). New York, NY: Springer.

Howley, L. D., & Martindale, J. (2004). The efficacy of standardized patient feedback in clinical teaching: A mixed methods analysis. *Medical Education, 9*(18), 1–10. doi: 10.3402/meo.v9i.4356

Intelligent Hearing Systems. (2016). I*ntroducing Baby Isao: The intelligent simulator for AEP and OAE* [Brochure]. Retrieved from http://www.ihsys.com/Brochures/BROCHURE_Simulator-Isao.pdf

Interprofessional Education Collaborative Expert Panel. (2011). *Core competencies for interprofessional collaborative practice: Report of an expert panel.* Washington, DC: Interprofessional Education Collaborative. Retrieved from http://www.aacn.nche.edu/education-resources/ipecreport.pdf

Issenberg, S. B., & Scalese, R. J. (2008). Simulation in health care education. *Perspectives in Biology and Medicine, 51*(1), 31–46. doi: 10.3109/0142159X.2013.818632

Kaf, W. A., Masterson, C. G., Dion, N., Berg, S. L., & Abdelhakiem, M. K. (2013). Optimizing otoscopy competency in audiology students through supplementary otoscopy training. *Journal of American Academy of Audiology, 24*, 859–866. doi: 10.3766/jaaa.24.9.9

Karakus, A., Duran, L., Yavuz, Y., Altintop, L., & Caliskan F. (2014). Computer based simulation training in emergency medicine designed in the light of malpractice cases. *BMC Medical Education, 14*, 155. doi: 10.1186/1472-6920-14-155

Kim, S., Phillips, W. R., Pinsky, L., Brock, D., Phillips, K., & Keary, J. (2006). A conceptual framework for developing teaching cases: A review and synthesis of the literature across disciplines. *Medical Education, 40*(9), 867–876. doi: 10.1111/j.1365-2929.2006.02544.x

Knudsen, L., Laplante-Levesque, A., Jones, L., Preminger, J. E., Nielsen, C., Lunner, T., ... Kramer, S. E. (2012). Conducting qualitative research in audiology: A tutorial. *International Journal of Audiology, 51*, 83–92. doi: 10.3109/14992027.2011.606283

Lapkin, S., & Levett-Jones, T. (2011). A cost-utility analysis of medium vs. high-fidelity human patient simulation manikins in nursing education. *Journal of Clinical Nursing, 20*, 3543–3552. doi: 10.1111/j.1365-2702.2011.03843.x

Lederman L. C. (1992). Debriefing: Toward a systematic assessment of theory and practice. *Simulation and Gaming, 23*(2), 145–160. doi: 10.1177/1046878192232003

Lestander, Ö., Lehto, N., & Engström, Å. (2016). Nursing students' perceptions of learning after high fidelity simulation: Effects of a three-step post-simulation reflection model. *Nurse Education Today, 40*, 219–224. doi: 10.1016/j.nedt.2016.03.011

Levett-Jones, T., McCoy, M., Lapkin, S., Noble, D., Hoffman, K., Dempsey, J., ... Roche, J. (2011). The development and psychometric testing of the Satisfaction with Simulation Experience Scale. *Nurse Education Today, 31*(7), 705–710. doi: 10.1016/j.nedt.2011.01.004

Lusk, J. M., & Fater, K. (2013). Postsimulation debriefing to maximize clinical judgment development. *Nurse Educator, 38*(1), 16–19. doi: 10.1097/NNE.0b013e318276df8b

Maran, N. J., & Glavin, R. J. (2003). Low- to high-fidelity simulation - A continuum of medical education? *Medical Education, 37*, 22–28. doi: 10.1046/j.1365-2923.37.s1.9.x

Meakim, C., Boese, T., Decker, S., Franklin, A. E., Gloe, D., Lioce, L., ... Borum, J. C. (2013). Standards of best practice: Simulation standard I: Terminology. *Clinical Simulation in Nursing, 9*, S3–S11. doi: 10.1016/j.ecns.2013.04.001

Morey J. C., Simon, R., Jay, G. D., Wears, R. L., Salisbury, M., & Dukes, K. A. (2003). Error reduction and performance improvement in the emergency department through formal teamwork training: Evaluation results of the Med Teams project. *Health Services Research, 37*, 1553–1581. doi: 10.1111/1475-6773.01104

Morgan, P. J., Tarshis, J., LeBlanc, V., Cleave-Hogg, D., DeSousa, S., Haley, M. F., ... Law, J. A. (2009). Efficacy of high-fidelity simulation debriefing on the performance of practicing anaesthetists in simulated scenarios. *British Journal of Anaesthesia, 103*(4), 531–537. doi: 10.1093/bja/aep222

Munoz, K., & Jeris, L. (2005). Learning to be interdisciplinary: An action research approach to boundary spanning. *Health Education Journal, 64*(5), 5–12. doi: 10.1177/001789690506400102

National Center for Hearing Assessment and Management. (2015). *Newborn hearing screening training curriculum (NHSTC): Interactive web based newborn hearing screening training curriculum*. Retreived from http://www.infanthearing.org/nhstc/index.html

Neill, M. A., & Wotton, K. (2011). High-fidelity simulation debriefing in nursing education: A literature review. *Clinical Simulation in Nursing, 7*(5), 161–168. doi: 10.1016/j.ecns.2011.02.001

Ng, S. L., Bartlett, D., & Lucy, S. D. (2012). Reflection as a tool for audiology student and novice practitioner learning, development and self-care. *Seminars in Hearing, 33*(2), 163–176. doi: 10.1055/s-0032-1311676

Ohtake, P. J., Marchilene, L., Schillo, R., & Rosen, M. (2013). Simulation experience enhances physical therapist student confidence in managing a patient in the critical care environment. *Physical Therapy, 93*(2), 216–228. doi: 10.2522/ptj.20110463

Owen, H., & Follows, V. (2006). GREAT simulation debriefing. *Medical Education, 40*, 488–489. doi: 10.1111/j.1365-2929.2006.02421.x

Passiment, M., Sacks, H., & Huang, G. (2011). *Medical simulation in medical education: Results of an AAMC survey*. Washington, DC: Association of American Medical Colleges. Retrieved from https://www.aamc.org/download/259760/data

Petranek, C. F. (2000). Written debriefing: The next vital step in learning with simulations. *Simulation & Gaming, 31*(1), 108–118. doi: 10.1177/104687810003100111

Pike, G. (1991). The effects of background, coursework, and involvement on students' grades and satisfaction. *Research in Higher Education, 32*(1), 15–31.

QSR International Pty Ltd. (2015). *NVivo qualitative data analysis software (Version 11)*. Retrieved from http://www.qsrinternational.com/nvivo-product

Rall, M., Manser, T., & Howard, S. (2000). Key elements of debriefing for simulator training. *European Journal of Anaesthesiology, 17*, 516–517. doi: 10.1046/j.1365-2346.2000.00724-1.x

Ryoo, E. N., & Ha, E-H. (2015). The importance of debriefing in simulation-based learning: Comparison between debriefing and no debriefing. *Computers, Informatics, Nursing, 33*(12), 538–545. doi: 10.1097/CIN.0000000000000194

Seropian, M. A. (2003). General concepts in full scale simulation: Getting started. *Anesthesia & Analgesia, 97*, 1695–1705. doi: 10.1213/01.ANE.0000090152.91261.D9

Shinnick, M. A., Woo, M., Horwich, T. B., & Steadman R. (2011). Debriefing: The most important component in simulation. *Clinical Simulation in Nursing, 7*(3), 105–111. doi: 10.1016/j.ecns.2010.11.005

Stillman, P. L., Swanson, D. B., Smee, S., Stillman, A. E., Ebert, T. H., Emmel, V. S., ... Willms, J. (1986). Assessing clinical skills of residents with standardized patients. *Annals of Internal Medicine, 105*(5), 762–771. doi: 10.7326/0003-4819-105-5-762

Stroud, S. D., Smith, C.A., Edlund, B. J., & Erkel, E. A. (1999). Evaluating clinical decision-making skills of nurse practitioner students. *Clinical Excellence for Nurse Practitioners, 3*, 230–237.

Wotton, K., Davis, J., Button, D., & Kelton, M. (2010). Third-year undergraduate nursing students' perceptions of high-fidelity simulation. *Journal of Nursing Education, 49*(11), 632–639. doi: 10.3928/01484834-20100831-01

Wu, R., & Shea, C. (2009). Using simulations to prepare OT students for ICU practice. *Education Special Interest Section Quarterly, 19*(4), 1–4.

Yule, S., Flin, R., Paterson-Brown, S., & Maran, N. (2006). Non-technical skills for surgeons in the operating room: A review of the literature. *Surgery, 139*(2), 140–149. doi: 10.1016/j.surg.2005.06.017.

Acknowledgements

The authors would like to thank the UAMS Simulation Center personnel, audiology faculty, standardized parents, and student participants for their time and cooperation.

Conflicts of Interest

The authors declare no potential conflicts of interest.

Authors' Note

Correspondence concerning this article should be addressed to Ahmad A. Alanazi, Department of Audiology and Speech Pathology, College of Applied Medical Sciences, King Saud bin Abdulaziz University for Health Sciences (KSAU-HS), Riyadh, Saudi Arabia. Email: alanaziahm@ksau-hs.edu.sa.

Appendix A
Students' Perception of Simulation Training Components (SPSTC) Survey

Event: _____ Student's Number _____ Date: _____

Please indicate your level of agreement (satisfaction) with each statement

The Simulation Training		Strongly Disagree (Very Dissatisfied)	Disagree	Mostly Disagree	Somewhat Agree	Mostly Agree	Agree	Strongly Agree (Very Satisfied)	Not Applicable
		1	2	3	4	5	6	7	N/A
Simulation Environment	1. The orientation to simulation was suitable								
	2. The simulation center was well equipped								
	3. The simulation environment felt safe for participation								
Learning Experience	4. The case scenarios were realistic								
	5. The debriefing sessions were well prepared								
	6. The debriefing sessions helped me reflect on my practice								
	7. The facilitator was supportive								
	8. Standardized parents acted as real parents								
	9. Standardized parents provided useful feedback								
	10. The learning experience will help me in my clinical practice								
Please respond to the following statements									
Simulation Training Highlights	11. Describe any part of the simulation training that was exceptional								
	12. Describe any part of the simulation training that did not meet your expectations								

Diagnosing Apraxia of Speech on the Basis of Eight Distinctive Signs

Diagnostiquer l'apraxie de la parole en se basant sur huit signes distinctifs

KEY WORDS
APRAXIA OF SPEECH
DIAGNOSIS
SPECIFIC SIGNS

Roel Jonkers*
Judith Feiken*
Ilse Stuive

*The first two authors contributed equally to the manuscript.

Abstract

This paper reports the results of a study on the use of a fixed number of specific signs to differentially diagnose Apraxia of Speech (AoS) from aphasia or dysarthria. This was done with a diagnostic instrument for AoS that was developed in the Netherlands in 2012, the Diagnostic Instrument for Apraxia of Speech (DIAS; Feiken & Jonkers, 2012). There were 8 signs identified as specific to AoS, namely: inconsistency of errors, number of errors with consonants versus vowels, difference between sequencing and alternating diadochokinesis, groping, initiation problems, syllable segmentation, cluster segmentation, and articulatory complexity. The DIAS was administered to 30 individuals with AoS, 10 individuals with aphasia, 10 individuals with dysarthria, and 35 control individuals. Results showed that a differential diagnosis could be made in 88% of the cases using a minimum of 3 out of 8 specific signs of AoS as criteria. With the exception of 2 patients with aphasia, no other group exhibited the presence of 3 or more signs of AoS. It was concluded that the presence of 3 signs is sufficient to differentially diagnose AoS from aphasia and dysarthria, despite the fact that there is a large amount of variability in the presence of signs of AoS itself in the different individuals.

Roel Jonkers
University of Groningen,
Groningen,
THE NETHERLANDS

Judith Feiken and
Ilse Stuive
Center for Rehabilitation,
University Medical Center,
Groningen,
THE NETHERLANDS
University of Groningen,
Groningen,
THE NETHERLANDS

Abrégé

Cet article présente les résultats d'une étude investiguant l'utilisation d'un nombre spécifique de signes pour distinguer l'apraxie de la parole de l'aphasie ou de la dysarthrie dans un processus de diagnostic différentiel. Pour ce faire, un test d'évaluation de l'apraxie de la parole ayant été développé aux Pays-Bas en 2012, soit le *Diagnostic Instrument for Apraxia of Speech* (DIAS; Feiken et Jonkers, 2012), a été utilisé. Huit signes ont été identifiés comme étant spécifiques à l'apraxie de la parole : inconstance des erreurs, nombre d'erreurs sur les consonnes versus les voyelles, différence entre les séries diadococinésiques en séquence et en alternance, tâtonnement, problèmes d'initiation, segmentation des syllabes, segmentation des groupes consonantiques et complexité articulatoire. Le *DIAS* a été administré à 30 participants ayant une apraxie de la parole, 10 participants ayant une aphasie, 10 participants ayant une dysarthrie et 35 participants formant un groupe contrôle. Les résultats ont montré qu'un diagnostic différentiel de l'apraxie de la parole peut être effectué dans 88% des cas en utilisant un minimum de trois critères sur huit. Aucun participant inclus dans les autres groupes expérimentaux n'a été identifié avec un minimum de trois signes spécifiques à l'apraxie de la parole, à l'exception de deux participants ayant une aphasie. La présence de trois signes spécifiques a ainsi été jugée suffisant pour distinguer l'apraxie de la parole de l'aphasie ou de la dysarthrie, et ce, malgré le fait qu'il existe une grande variabilité dans les signes observés au sein des individus ayant une apraxie de la parole.

The importance of standardizing the assessment of Apraxia of Speech (AoS) has been repeatedly emphasized in scientific literature (Knollman-Porter, 2008; Wambaugh, 2006; West, Hesketh, Vail, & Bowen, 2008; World Health Organization, 2005). AoS is generally defined as an impairment in programming the positioning of speech organs and the sequencing of articulations (Darley, 1968; Ziegler, 2008). There is, however, no consensus on how to diagnostically differentiate AoS from related communication disorders such as aphasia and dysarthria (Ziegler, Aichert, & Staiger, 2012). Also, there is still a debate in scientific circles regarding which particular signs lead to the diagnosis of AoS (Lowit, Miller, & Kuschmann, 2014; McNeil, Pratt, & Fossett, 2004; Ziegler, 2008).

To diagnose AoS, in the Netherlands, speech-language pathologists (S-LPs) usually administer general language tests or a dysarthria test (Feiken, Hofstede, & Jonkers, 2008; Jonkers, Terband, & Maassen, 2014), or base their diagnosis on clinical judgments. Internationally, there are a few standardized and normed instruments available, like the Apraxia Battery for Adults (ABA-2; Dabul, 2000) and the Motor Speech Examination (MSE; Ogar et al., 2006; Wertz, LaPointe, & Rosenbek, 1984) for English, as well as the *Hierarchische Wortlisten* (Liepold, Ziegler, & Brendel, 2002) for German. There are also criteria lists available to identify AoS, such as the Mayo Clinic Apraxia of Speech Battery (Darley, Aronson, & Brown, 1975; Duffy, 2005; Wertz et al., 1984); the checklist of McNeil, Robin, and Schmidt (2009); and the Academy of Neurologic Communication Disorders and Sciences (ANCDS) list (Knollman-Porter, 2008; Wambaugh, 2006). However, according to Knollman-Porter (2008) and West et al. (2008), there are no instruments or lists that provide reliable identification of AoS.

A recently developed tool, published by Strand, Duffy, Clark, and Josephs (2014), could be valuable in diagnosing (progressive) AoS. Strand et al. presented a rating scale for the diagnosis and description of AoS and tested this in a group of participants with (progressive) AoS or aphasia, reporting high reliability scores as well as good validity of the tool. In the same vein, the current study investigates whether the identification of specific signs is useful for the differential diagnosis of AoS. These signs were measured with a recently developed Dutch diagnostic test, the Diagnostic Instrument for Apraxia of Speech (DIAS; Feiken & Jonkers, 2012). In contrast to the scale tool of Strand et al. (2014), the DIAS could be valuable in differentially diagnosing stroke-induced AoS from dysarthria and aphasia. The diagnosis is based on the presence of eight signs indicative of AoS, which were carefully selected based on theories regarding the characteristics and nature of AoS.

A literature review revealed 33 distinctive signs of AoS, which were categorized by the authors into primary and secondary signs. The primary signs were categorized into three subgroups: initiation errors, incorrectly articulated phonemes, and sequencing errors. Initiation errors include pausing before an utterance, visible or audible struggle to position the articulators (groping), and restarts (Duffy, 2005; Strand et al., 2014). Incorrectly articulated phonemes lead to signs like distortions or substitutions. In distortions, the target phoneme is still recognizable. If it is no longer possible to recognize the target phoneme a substitution occurs, where the change of one or more features leads to the production of another phoneme (den Ouden, 2002). Sequencing errors are exchange errors at the level of sound or syllable (Haynes, Pindzola, & Emerick, 1992; Square, Roy, & Martin, 1997; Ziegler, 2008). The number of initiation errors and distortions seems to be affected by articulatory complexity as well (Staiger & Ziegler, 2008). This is reflected at the phoneme level in a larger number of errors with consonants as compared to vowels (Wertz et al., 1984), and at the word level in a larger number of errors with syllables containing consonant clusters as compared to simple syllables (Staiger & Ziegler, 2008).

The secondary signs are signs that can be assumed to be reactions to the underlying disorder. Speakers with AoS may pause more often (Aichert & Ziegler, 2004; Duffy, 2005) between the consonants of a cluster (cluster segmentation; McNeil, 2002) or the syllables of a word (syllable segmentation; Staiger & Ziegler, 2008), and lengthen vowels (Van der Merwe, 2009). In so doing, individuals with AoS create more time for articulatory motor programming, to lower the number of articulation errors.

The categorized primary and secondary signs were compared to the signs seen in other neurologic speech disorders, like aphasia and dysarthria. Overlapping signs were omitted. Examples of these signs are a word-length effect or the presence of substitutions, which are signs that can be found in both AoS and aphasia (Romani & Galluzzi, 2005; Ziegler, 2005). A sign that is found in both individuals with AoS and dysarthria is slow speech, but also problems with diadochokinesis in general (Duffy, 2005; Ziegler, 2002). However, as alternating diadochokinesis (/pa-ta-ka/) is specifically more difficult for individuals with AoS than sequencing diadochokinesis (/pa-pa-pa/; Ziegler, 2002), this characteristic can be considered as a specific sign of AoS.

The resulting eight signs were considered to be critical signs useful for the differential diagnosis of AoS. The scored signs are: 1) inconsistency in the pronunciation of repeated phonemes, 2) more errors with consonants than with vowels, 3) more difficulty in alternating diadochokinetic rate (/pa-ta-ka/) than sequencing diadochokenitic rate (/pa-pa-pa/), 4) visual or audible groping, 5) initiation problems (restarts), 6) syllable segmentation, 7) segmentation of consonant combinations, and 8) effect of articulatory complexity. These signs are assessed using the DIAS, which is described in the Methods section.

In this study, it will first be investigated whether the eight signs of AoS can be scored reliably. The diagnosis of AoS will be based on the presence of a number of these signs. For the differential diagnosis of AoS with aphasia and dysarthria not all signs need to be present, as the same underlying deficit may lead to different primary and secondary signs. How many signs need to be present in order to come to a differential diagnosis will be investigated. The outcomes of a study with 50 brain-damaged speakers and 35 non-brain-damaged control speakers will be presented.

Methods

Participants

Participants were selected as possibly having AoS by the treating S-LP based on the most recent criteria, i.e., the ANCDS list (Wambaugh, 2006). Another S-LP then independently confirmed this judgment. This S-LP was blinded to the diagnosis of the first S-LP. Both S-LPs were independent in the sense that they were not co-authors of the article. This study only considered those cases where both S-LPs agreed on the clinical diagnosis of AoS.

Thirty participants (15 male, 15 female; mean age 58.4 years, range 34–78) clinically diagnosed with AoS were assessed with the DIAS. To study the potential of differentially diagnosing between patients with AoS, aphasia, and dysarthria on the basis of clinical signs, 20 participants without AoS but with aphasia (n = 10; eight male, two female; mean age 62.7 years, range 45–77) or dysarthria (n = 10; nine male, one female; mean age 55.8 years, range 18–77) were also tested with the DIAS. All participants with AoS suffered from a single stroke. The same holds for eight of the participants with dysarthria and eight of the participants with aphasia. One individual with aphasia and one with dysarthria suffered from a traumatic brain injury. One other individual with dysarthria suffered from a subarachnoid bleed, and for one participant with aphasia the specific etiology was unknown. Aphasia was diagnosed with the standard Dutch diagnostic test, the Aachen Aphasia Test (Graetz, De Bleser, & Willmes, 1992). Only participants having aphasia with phonological deficits, reflected in low scores for repetition and phonological errors in spontaneous speech, were included. Dysarthria was diagnosed with the Dutch *Radboud Dysartrie Onderzoek* [Radboud Dysarthria Investigation] (RDO; Knuijt & de Swart, 2007). The 20 participants without AoS were selected on the basis of their entry in the rehabilitation centre where this study was performed. The first 10 participants with aphasia and dysarthria—irrespective of the type of aphasia or dysarthria—that fit the inclusion criteria were tested. Therefore, this group was less balanced with respect to sex than the group with AoS. A group of 35 control speakers that matched the participants with AoS in age, sex, and education was also tested to determine the cut-off points for the different signs. This group consisted of 14 male and 21 female participants, mean age 52.3 years (range 23–64). A chi-square test revealed no difference between the AoS group and the control group with respect to sex ($\chi(1) = 0.754$, $p > .05$). However, the AoS group turned out to be significantly older than the control group ($t(63) = 2.489$, $p < .05$). Nevertheless, the mean age of both groups was below 60, and in the AoS group only four of the 35 participants were older than 70. Therefore, age is not assumed to be of influence on the outcomes.

All participants gave their informed consent. Testing was done with permission of the Medical Ethics Committee of the University Hospital Groningen (UMCG). All participants were native speakers of Dutch. Participants had a normal intellect (IQ > 70) and vision, and their hearing and neurocognitive abilities did not interfere with an acceptable assessment. All patient group data are presented in Table 1.

All individuals with AoS also suffered from aphasia. In order to determine if the results of this study could be explained by a difference in the severity of aphasia between the group with AoS and aphasia, their scores on the Token Test of the AAT were compared. Originally, the Token Test was developed to be a test for the reception of language, but currently the Token Test is used as a selective instrument to detect the presence of aphasia and as an indicator of its severity (El Hachioui et al., 2013; Orgass & Poeck, 1966). The maximum score on this test is 50, which reflects a negative score. Individuals without aphasia had a mean score of 2.4 (SD = 2.5) on this test (Graetz et al., 1992).

Table 1. Participants by Speech Category

Group	Age in years (Mean and *SD*)	Sex	TPO in months (Mean and *SD*)
Apraxia of speech (*n* = 30)	58.4 (11.6)	15 m, 15 f	32.0 (25.4)
Aphasia (*n* = 10)	62.7 (9.8)	8 m, 2 f	29.7 (53.9)
Dysarthria (*n* = 10)	55.8 (16.3)	9 m, 1 f	10.5 (4.4)
Control speakers (*n* = 35)	52.3 (11.3)	14 m, 21 f	-

Note. m = male, f = female, TPO = time post onset

Materials

All individuals were tested with the DIAS (Feiken & Jonkers, 2012). The DIAS contains four tests, of which three were used in this study.[1] The test for orofacial apraxia will not be discussed here, as it is only part of the instrument to diagnose orofacial apraxia. Three tests were administered to assess the presence of the eight aforementioned signs: articulation of phonemes, diadochokinesis, and articulation of words. All items can be found in Appendix A. In Table 2, an overview is given of the three tasks that were used for differential diagnosis, mentioning the different signs that were studied. Not all signs were investigated in every subtest, but the three subtests were indicated for specific signs. Below, the tests are described including descriptions of the specific signs per test.

In the test for the articulation of phonemes, participants are instructed to repeat vowels and consonants three times consecutively. This test evaluates the conscious production of individual consonants. In AoS, inconsistent distortions and substitutions of phonemes often occur (Sign 1; Darley et al., 1975; den Ouden, 2002; Varley & Whiteside, 2001;Wertz et al., 1984). Inconsistent errors in this study are assumed to be different pronunciations during the repetition of three phonemes. Wambaugh (2006) states that errors of speakers with AoS are consistent. However, this is a different kind of consistency, as it refers to the consistency of error types across different tests.

With respect to the number of errors made with consonants or vowels (Sign 2), more errors with

Table 2. Subtests of the DIAS

Test	Differential diagnostic criteria	Control score mean (*SD*)	Cut-off score
Articulation of phonemes (15 consonants; 15 vowels)	- Inconsistency of errors (1) - Number of errors with consonants vs. vowels (2)	0.09 (0.51) 0.09 (0.74)	2 2
Diadochokinesis (6 series of sequencing and alternating syllables or words)	- Difference between sequencing and alternating diadochokinesis (3) - Groping (4)	0.94 (0.11) 0	0.74 2*
Articulation of words (6 blocks of 11 words)	- Initiation problems (5) - Syllable segmentation (6) - Cluster segmentation (7) - Articulatory complexity (8)	0.003 (0.02) 0 0 0.10 (0.39)	1 out of 11 blocks > 0 > 0 0.88

Note. *Groping was not seen in the control group, thus every occurrence could be considered deviant. However, as clinicians questioned this symptom during the pilot phase on certain occasions, the cut-off was set to 2.

[1]It is not intended to provide an elaborate description of the subtests and the theoretical background of the DIAS. Feiken and Jonkers (2012) and Jonkers et al. (2014) provide more information on construct and item validity, specificity, and sensitivity of the test.

consonants than with vowels are expected (Duffy, 2005; Wertz et al., 1984).

The test for the articulation of phonemes consists of 30 items: 15 consonants (C) and 15 vowels (V). This composition allows one to assess whether there is a difference in the number of errors between consonants and vowels. Consonants differed in place or manner of articulation. Vowels were chosen on the basis of their position in the vowel triangle (Kooij & van Oostendorp, 2003). Place of articulation of the consonants was varied to circumvent perseveration. After, for example, the consonant /m/, an alveolar sound like /d/ followed. The internal consistency of this test is .96 (Cronbach's alpha). To account for a possible effect of consistency, participants were asked to repeat every phoneme three times in a row.

The second test in the DIAS that plays a role in differential diagnosis is a diadochokinesis task. Oral diadochokinesis is seen as a sensitive measure for neuromotoric speech capacities (Ziegler, 2002), as it demands maximum performance of a participant. Deger and Ziegler (2002), Ogar et al. (2006), and Wertz et al. (1984) note that individuals with AoS will have more difficulties in alternating different syllables (alternating diadochokinesis) than repeating the same syllables (sequential diadochokinesis), which is defined as Sign 3.[2] Initiation problems, substitutions, omissions, slow speech rate, segmentation of syllables or clusters, and repeated attempts to produce an item are possible consequences of difficulties with alternating diadochokinesis. In accordance with Duffy (2005), the diadochokinesis test was also specifically used to observe the symptom of groping (Sign 4).

The diadochokinesis test contains 12 items: six sequencing and six alternating items. This subtest is set up according to the level of complexity, starting with simple CV structures, like the sequencing item /pa-pa-pa/ versus the alternating item /pa-ta-ka/, and ending with CCVCC structures, like /stank-stank-stank/ versus /stank-blank-drank/. In some of the alternating items the consonant in initial or final position changes, whereas in others the consonants within a cluster change. Most of the words used in these structures were meaningful words. The words were controlled for frequency of occurrence using the CELEX frequency list for Dutch (Baayen, Piepenbrock, & Gulikers, 1995). The sequential items always had the lowest frequency, to prevent any poor performance on the alternate version of the item, which could be explained by a word frequency effect. The internal consistency of this test is .97 (Cronbach's alpha).

With the test for the articulation of words, the presence of the final four signs of AoS is studied, among which are initiation problems (Sign 5). Problems with the initiation of speech are often seen in individuals with AoS (Haynes et al., 1992; LaPointe, 1990). They can appear in different forms. LaPointe (1990) describes false starts and repetition of sounds or syllables as instances of initiation problems. As mentioned in the introduction, as a reaction to articulation problems speakers with AoS may also pause more often, leading to cluster segmentation (McNeil, 2002; Sign 6) or syllable segmentation (Staiger & Ziegler, 2008; Sign 7). Finally, individuals with AoS make more repetition errors with consonant clusters (Staiger & Ziegler, 2008) and with longer words (Ziegler, 2005), and this is reflected in the *articulatory complexity* sign (Sign 8).

The test for the articulation of words (word repetition) contains 66 items with increasing length and articulatory complexity. The test consists of 11 blocks of six words, where every block differed in complexity, with respect to the number of syllables, number of phonemes and articulatory complexity (CV structures, CC clusters within a syllable, CCC clusters within a syllable, and CC clusters at the syllable boundary). Every block of six items focused on a specific structure. The words in the test do not differ with respect to word frequency. The internal consistency of this test is .99 (Cronbach's alpha). Kuschmann, Miller, and Lowit (2014) provide requirements for intelligibility tests used in speakers with AoS, considering, among others, adequacy, completeness, levels of difficulty, number of items, and frequency of items. The list of items in this test fits with the requirements mentioned here.

Procedure

All tests were administered in one session in a fixed order. All assessments were videotaped and scored later. The administration of the subtests was multimodal, meaning that the items were presented both visually and auditorily to circumvent influences of visual or auditory problems. Participant and tester sat face-to-face in a quiet room. To prevent lip reading, the participant was asked not to look at the tester during the assessment. In cases where this was not possible, the mouth of the tester was covered. Testing (including the test for orofacial apraxia) lasted about 45 minutes. After instruction, all subtests

[2]There is some confusion in the literature about what should be seen as sequential diadochokinesis and what should be seen as alternating diadochokinesis. Duffy (2005), for example, uses the terms with the inverse meaning. However, there is agreement on the fact that the repetition of different syllables, like /pa-ta-ka/, is more difficult for individuals with AoS than the repetition of the same syllable (/pa-pa-pa/).

started with two examples. In the case of an inadequate response to (one of the) examples, the participants were corrected. During assessment no help or feedback was provided, except for one repetition of the target if the participant requested it. There was no time pressure to answer, except for in the diadochokinesis test. In this test participants were first asked to repeat every sequence of three syllables once, and if this was possible, they were asked to repeat every sequence as often and correctly as possible within 8 seconds. The tester told the participant when to start and stop.

Scoring

For each test, the presence of the specific signs was evaluated. Cut-off points for the presence of signs were determined based on scores of the control speakers. A symptom was considered to be present if a score differed more than two standard deviations from the mean score of the control speakers (adjusted upwards if necessary). These cut-off points are presented in Table 2. In Appendix B, how the specific signs were scored per test is described. Scoring and interpreting of the errors could be done in 45 minutes.

The number of signs was counted for every participant and it was evaluated whether it was possible to distinguish individuals with AoS from individuals with dysarthria or aphasia based on the number of signs present. Severity is not considered in the current study. This means that the presence of a sign is important but the frequency with which a sign is noted is not.

Reliability

Intra-rater reliability was obtained by comparing the scoring of sign presence on the basis of video recordings of the DIAS of 30 participants twice, with an intermediate period of six months, by the same experienced clinical linguist. Inter-rater reliability was based on the scores of three experienced S-LPs not involved in the intra-rater reliability, who scored the video recordings of the DIAS administration independently. Test-retest reliability was obtained by testing 10 participants with the DIAS twice, with an intermediate period between two and six weeks. Again, video recordings were scored.

Results

Reliability

Intra-rater and inter-rater reliability correlations (intra-class correlations (ICC) or Kappa scores) for scoring the eight signs were significant and showed overall good reliability. In Table 3, the ICC values and Kappa scores for all reliability measures are presented. The lowest inter-rater agreement was seen for *cluster segmentation*, although this agreement is still acceptable. There was a strong agreement for *more errors with consonants than with vowels*. All other intra-class correlations showed very high agreement. The Kappa values for groping indicated good to excellent agreement. With respect to the ICC values for the intra-rater reliability, all correlations were significant at the level of .001 and indicated a very high agreement. Not all correlations were significant for the test-retest reliability. A non-significant and poor agreement was found for *articulatory complexity*. Ratings for the other signs were again significant, and agreement varied from good (groping) or strong (cluster segmentation) to very high (all other signs).

Number of signs

In order to find out what the necessary number of signs would be for the diagnosis of AoS, the number of signs in the three groups was calculated and afterwards it was decided what the ideal number needed for a reliable diagnosis would be. In comparing the presence of signs in participants with AoS with those noted in individuals with dysarthria and aphasia, it was found that the presence of at least three signs was needed to diagnose AoS in most of the individuals with AoS. In 26 of the 30 individuals with AoS, three or more signs were determined. Three of the four individuals with fewer signs were individuals with very severe speech problems. In these individuals only the first two subtests could be administered, and therefore most of the signs could not be determined. Only in one case a participant was able to do all the subtests and still had fewer than three signs. Three individuals with AoS, however, showed only three signs, which means that when using four signs as diagnostic criteria, a smaller number of individuals with AoS would be diagnosed properly.

In the group of individuals with dysarthria (n = 10), none of the individuals had three or more signs of AoS. In the aphasia group (n = 10), two individuals had three signs of AoS, while the other eight individuals showed fewer signs. Seven individuals not assumed to have AoS showed two signs, which would lead to a larger number of misdiagnoses if these were to be used as diagnostic criteria. This means that the presence of three signs was the best way to divide the groups into individuals with and without AoS.

In Table 4, an overview is provided with the number of individuals in the AoS group that displayed a specific sign. Every sign was found in almost half of the speakers

Table 3. Reliability Measures for the Different Signs

	Inter-rater reliability	Intra-rater reliability	Test-retest reliability
Inconsistent realization of phonemes	.84 ($p < .01$)	.98 ($p < .001$)	.93 ($p < .001$)
More errors with consonants than with vowels	.76 ($p < .01$)	.95 ($p < .001$)	.81 ($p < .05$)
More problems with alternating than with sequencing syllables	.92 ($p < .001$)	.98 ($p < .001$)	.98 ($p < .001$)
Initiation problems	.81 ($p < .001$)	.95 ($p < .001$)	.92 ($p < .001$)
Syllable segmentation	.81 ($p < .001$)	.98 ($p < .001$)	.99 ($p < .001$)
Cluster segmentation	.62 ($p < .001$)	.90 ($p < .001$)	.73 ($p < .05$)
Articulatory complexity	.80 ($p < .001$)	.95 ($p < .001$)	.32 ($p > .05$)
Groping	Kappa		
Rater 1-2	.73 ($p < .05$)	.86 ($p < .001$)	.74 ($p < .05$)
Rater 1-3	.73 ($p < .05$)		
Rater 2-3	1.00 ($p < .001$)		

Note. All comparisons: intra-class reliability, except for the *groping* sign, for which Kappa-scores were used.

Table 4. Number of Individuals With AoS Showing Specific Symptoms of AoS

Signs	Individuals with AoS Symptoms ($n = 30$)
Inconsistency of errors	17/30
Number of errors with consonants vs. vowels	13/30
Difference between sequencing and alternating diadochokinesis	18/30
Groping	23/30
Initiation problems	28/30
Syllable segmentation	25/30
Cluster segmentation	14/30
Articulatory complexity	18/30

with AoS. The sign *more errors with consonants than with vowels* was found in the lowest number of speakers with AoS. Only 13 of the 30 speakers showed this sign. In almost all speakers with AoS (28/30), initiation errors occurred. No specific pattern was seen with respect to the number of primary or secondary signs.

Severity of aphasia

The mean Token Test score of the individuals with AoS was 24.9 (*SD* = 13.5) and of the individuals with only aphasia 27.0 (*SD* = 16.6). An unpaired *t*-test did not show a significant difference between these scores (t(38) = 0.4, $p > .05$). This indicates that differences between the groups with respect to the presence of signs do not relate to the severity of the aphasia.

Discussion

The current study investigated whether it is possible to differentially diagnose AoS from dysarthria or aphasia on the basis of the presence of signs of AoS. With the Dutch DIAS (Feiken & Jonkers, 2012), the presence of eight specific signs of AoS was studied in a group of individuals with AoS, dysarthria, and aphasia, as well as a control group. The individuals with AoS were selected on the basis of clinical judgment by an S-LP using the most recent selection criteria for AoS, i.e., the ANCDS list (Wambaugh, 2006) and this judgment was independently confirmed by the judgment of a second blinded S-LP. The individuals with dysarthria and aphasia were diagnosed with the RDO (Knuijt & de Swart, 2007) and the Aachen Aphasia Test (Graetz et al., 1992), respectively.

Haley, Jacks, De Riesthal, Abou-Khalil, and Roth (2012) showed that clinicians are reliably able to list and interpret the signs of AoS, but show poor agreement in differentially diagnosing AoS. This is because clinicians observe and prioritize the signs differently, and consequently reach different conclusions. In this study, we showed that the eight signs can be scored reliably by experienced S-LPs. Both the inter- and intra-rater reliability showed significant and sufficiently high correlations. This also holds for the test-retest variability, except for the *articulatory complexity* sign. Although this sign was found in 18 of the 30 speakers with AoS, it seems that the presence of this sign is not as clear to interpret as the other signs. This might have to do with the fact that the calculation of this sign is more complex than the other signs, although the inter- and intra-rater reliability were good. It could also be that the presence of this sign is subtler to detect than the others, which means that in some cases raters might miss its presence. For future studies it is recommended to detect the presence of the influence of articulatory complexity with a simpler measure.

In the individuals with AoS, the signs were present, but with a large amount of variation. This is consistent with the assumption that the same underlying disorder can manifest itself in different primary or secondary signs. However, the differential diagnosis could be determined with the presence of three of eight signs. In 26 of 30 tested individuals with AoS, three or more signs were present. Three of the four remaining individuals could not be diagnosed properly as they were severely impaired patients who could not complete all the subtests. These individuals were for example unable to do the diadochokinesis test at all, or could only repeat one or two words of the repetition test. Therefore, in these individuals not all signs could be counted. This leads to a restriction on a valid diagnosis on the basis of signs, namely that individuals should be assessed with the entire diagnostic test and that all signs can at least be scored properly. Only one individual with a clinical diagnosis of AoS scored with fewer than three signs. For this individual it is difficult to decide whether he/she was incorrectly diagnosed with AoS by the S-LP or incorrectly diagnosed as not having AoS using the DIAS.

The presence of three of more signs was not seen in any of the ten individuals with dysarthria; however, three signs were present in two of the 10 individuals with aphasia. This result can be interpreted in two different ways. One could conclude that it is not always possible to make a differential diagnosis between aphasia and AoS in some cases. Another possible interpretation is that diagnosing on the basis of the presence of symptoms is preferable to clinical judgment, because of the possibility that these aphasic speakers also suffer from AoS. Control speakers and the individuals in the other patient groups rarely showed these signs. The fact that more signs were present in the group of speakers with AoS than in the group of speakers with aphasia appeared to be unrelated with severity of aphasia, because Token Test scores for both groups were comparable. However, there is some debate about the role of the Token Test as a measure for severity of aphasia. Although authors use the Token Test in such a way (e.g., El Hachioui et al., 2013), the developers of the Token Test originally presented it as an instrument to diagnose language comprehension impairments only (see also De Renzi & Faglioni, 1978). In that case, the only justified conclusion is that the presence of three or more distinctive signs in participants with AoS in the current study seems unrelated to the presence of an aphasic comprehension disorder.

No specific signs seem to favour the diagnosis of AoS. All signs were found regularly in the different individuals, with a minimum of 13 out of 30 speakers with AoS showing the sign of more errors with consonants than with vowels. This reveals that it does not seem to be possible to further restrict the number of symptoms to be present. It is also clear that not all signs are found in all individuals with AoS. Only the symptom of initiation problems was seen in almost all speakers.

As mentioned in the introduction, there is a lively debate on the diagnosis of AoS and on which type of tasks to use for diagnosis. There is a discussion in the literature as to what importance non-speech tasks, such as repetition of phonemes and diadochokinesis, could contribute to the diagnosis of AoS. Ziegler (2003) doubts the role of such tasks due to their unrelatedness to natural speech. It is indeed impossible to diagnose AoS on the sole basis of such tasks, but in line with Kuschmann et al. (2014), it is assumed that these non-speech tasks provide information on the underlying impairment, whereas it is also necessary to focus on real words, as is done in the repetition task, for a closer correlation with natural speech. Both types of tasks, therefore, have merit in the assessment of AoS. The importance of a diadochokinesis test is also reflected in the fact that 18 of the individuals with AoS showed greater problems with alternating diadochokinesis as compared to sequential diadochokinesis.

The discussion on diagnosis partly has to do with the lack of consensus on the exact underlying deficit(s) and the differential diagnosis with respect to aphasia and dysarthria. There seems to be agreement on some of the signs of AoS, but even with respect to these signs there is discussion regarding whether they should really be seen as purely signs of AoS. All of the eight signs that are evaluated in the DIAS were mentioned as signs of AoS in the literature. The assertion that not all eight signs need to be present in all individuals with AoS has been shown in this study and was also confirmed by Strand et al. (2014). Strand and colleagues recently showed that it is possible to reliably score the presence of signs of AoS and to validly diagnose (progressive) AoS on the basis of the presence of these signs, also without the necessity of all signs being present for a group of individuals with (progressive) AoS (Strand et al., 2014).

In line with the findings of Strand et al. (2014), the current study indicates that the discussion about the differential diagnosis with respect to aphasia and dysarthria should not be about finding signs that are present in all AoS patients. When the division of the signs of AoS into primary signs (like initiation errors and distortions) and secondary signs (like segmentation of consonant clusters or intersyllabic pauses) is taken into account, it is likely that individuals differ in how they express AoS. Therefore, different specific signs could lead to the diagnosis of AoS. The present study showed that, nevertheless, only three signs need to be present to result in a valid differential diagnosis between speakers with and without AoS.

This current study is limited by the fact that, although a significant number of individuals with AoS participated, the groups of individuals with dysarthria and aphasia were rather small. Accordingly, no specific distribution was made in the different types of individuals with dysarthria (e.g., ataxic dysarthria or flaccid dysarthria) or aphasia (e.g., conduction aphasia or Wernicke's aphasia). In future studies, the authors intend to account for the type of dysarthria or aphasia by testing a larger number of participants.

A second limitation is the fact that this study was conducted with Dutch participants using a Dutch instrument. It is assumed, however, that the specific symptoms that were considered with this instrument might be considered in other languages as well. The fact that Strand et al. (2014) were able to use signs for the diagnosis of AoS shows that a diagnosis on the basis of the presence of signs does not have to be test-specific.

Finally, the fixed order of the subtests could have influenced the outcomes. Participants might have had more speech problems at the beginning of the administration of the test due to starting problems, or at the end due to, for example, fatigue, which could lead to a bias in the presence of specific symptoms. However, given the fact that no specific sign was the most common in the participants with AoS, it seems unlikely that more symptoms would be shown in the first or final test for the group of participants with AoS.

In this study, it was shown that, by assessing the specific signs of AoS, AoS can be distinguished from aphasia and dysarthria. The possibility of differentially diagnosing AoS from aphasia and dysarthria is important in clinical practice. S-LPs will be able to connect their treatment properly to the actual deficit(s), creating a better basis for treatment. In addition, by knowing which signs are present in a specific patient, better choices can be made in setting priorities for therapy. Administration of the test and scoring of the responses can be done in roughly 90 minutes. S-LPs are able to do the scoring

and interpretation on the basis of the description in the manual. One-day courses are also offered, however, to acquaint S-LPs with the procedures. Haley et al. (2012) already showed that S-LPs often have different opinions on the presence of a sign, but that training on the basis of a systematic protocol clearly reduces these differences.

References

Aichert, I., & Ziegler, W. (2004). Syllable frequency and syllable structure in apraxia of speech. *Brain and Language, 88*(1), 148–159. doi: 10.1016/S0093-934X(03)00296-7

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database (Release 2)* [CD-Rom]. Philadelphia, PA: Linguistic Data Consortium.

Dabul, B. (2000). *Apraxia battery for adults* (2nd ed). Austin, TX: Pro-Ed.

Darley, F. L. (1968). *Apraxia of speech: 107 years of terminological confusion*. Paper presented at the American Speech and Hearing Association Convention, Denver, CO.

Darley, F., Aronson, A., & Brown, J. (1975). *Motor speech disorders*. Philadelphia, PA: Saunders.

Deger, K., & Ziegler, W. (2002). Speech motor programming in apraxia of speech. *Journal of Phonetics, 30*, 321–335. doi: 10.1006/jpho.2001.0163

Den Ouden, D. (2002). *Phonology in aphasia*. Unpublished doctoral dissertation, Rijksuniversiteit Groningen, Groningen, The Netherlands.

De Renzi, E., & Faglioni, P. (1978). Normative data and screening power of a shortened version of the Token Test. *Cortex, 14*(1), 41–49.

Duffy, J. R. (2005). *Motor speech disorders: Substrates, differential diagnosis, and management*. St Louis, MO: Elsevier Mosby.

El Hachioui, H., Lingsma, H. F., van de Sandt-Koenderman, M. E., Dippel, D. W., Koudstaal, P. J., & Visch-Brink, E. G. (2013). Recovery of aphasia after stroke: A 1-year follow-up study. *Journal of Neurology, 260*(1), 166–171. doi: 10.1007/s00415-012-6607-2

Feiken, J., Hofstede, G., & Jonkers, R. (2008). De diagnostiek van verbale apraxia [The diagnosis of verbal apraxia]. *Logopedie en Foniatrie, 7/8*, 228-234.

Feiken, J., & Jonkers, R. (2012). *Diagnostisch instrument voor apraxie van de spraak* [DIAS; Diagnostic Instrument for Apraxia of Speech]. Houten, The Netherlands: Bohn, Stafleu en Van Loghum.

Graetz, P., De Bleser, R., & Willmes, K. (1992). *Akense Afasietest* [Aachen Aphasia Test]. Lisse, The Netherlands: Swets and Zeitlinger.

Haynes, W., Pindzola, R., & Emerick, L. (1992). *Diagnosis and evaluation in speech pathology*. Englewood-Cliffs, NJ: Prentice-Hall.

Haley, K. L., Jacks, A., De Riesthal, M., Abou-Khalil, R., & Roth, H. L. (2012). Toward a quantitative basis for assessment and diagnosis of apraxia of speech. *Journal of Speech, Language, and Hearing Research, 55*(5), 1502–1517. doi: 10.1044/1092-4388(2012/11-0318)

Jonkers, R., Terband, H., & Maassen, B. (2014). Diagnosis and therapy in adult acquired dysarthria and apraxia of speech in Dutch. In N. Miller & A. Lowit (Eds.), *Motor speech disorder. A cross-language perspective* (pp. 156–167). Bristol, UK: Multilingual Matters.

Knollman-Porter, K. (2008). Acquired apraxia of speech: A review. *Topics in Stroke Rehabilitation, 15*(5), 484–493. doi: 10.1310/tsr1505-484

Knuijt, S., & de Swart, B. J. M. (2007). *Handleiding 'Radboud Dysartrieonderzoek'* [Manual Radboud Dysarthria Study]. Nijmegen, The Netherlands: UMC St Radboud.

Kooij, J., & van Oostendorp, M. (2003). *Fonologie. Uitnodiging tot de klankleer van het Nederlands* [Phonology. Invitation to the phonology of Dutch]. Amsterdam, The Netherlands: University Press.

Kuschmann, A., Miller, N., & Lowit, A. (2014). Motor speech disorders: Issues in assessment and management. In N. Miller & A. Lowit (Eds.), *Motor speech disorder. A cross-language perspective* (pp. 41–57). Bristol, UK: Multilingual Matters.

LaPointe, L. (1990). Neurogenic disorders of speech. In G. Shames & E. Wiig (Eds.), *Human communication disorders* (pp. 495–530). Columbus, OH: Merrill Publishing Company.

Liepold, M., Ziegler, W., & Brendel, B. (2002). *Hierarchische Wortlisten. Ein Nasprechtest für die Sprechapraxiediagnostik* [Hierarchical lists of words. A repetition test for the diagnosis of apraxia of speech]. Dortmund, Germany: Borgmann.

Lowit, A., Miller, N., & Kuschmann, A. (2014). Motor speech disorders: What are they?. In N. Miller & A. Lowit (Eds.), *Motor speech disorder. A cross-language perspective* (pp. 29–40). Bristol, UK: Multilingual Matters.

McNeil, M. R. (2002). *Clinical characteristics of apraxia of speech: Model/behaviour coherence*. Proceedings of the 2002 Childhood Apraxia of Speech Research Symposium, Carlsbad, CA.

McNeil, M. R., Pratt, S. R., & Fossett, T. R. D. (2004). The differential diagnosis of apraxia of speech. In B. Maassen, R. D. Kent, H. F. M. Peters, P. H. M. M. van Lieshout, & W. Hulstijn (Eds.), *Speech motor control in normal and disordered speech* (pp. 389–413). Oxford, UK: Oxford Medical Publications.

McNeil, M. R., Robin, D. A., & Schmidt, R. A. (2009). Apraxia of speech. In M. R. McNeil (Ed.), *Clinical management of sensorimotor speech disorders* (pp. 249–268). New York, NY: Thieme.

Ogar, J., Willock, S., Baldo, J., Wilkins, D., Ludy, C., & Dronkers, N. (2006). Clinical and anatomical correlates of apraxia of speech. *Brain and Language, 97*(3), 343–350. doi: 10.1016/j.bandl.2006.01.008

Orgass, B., & Poeck, K. (1966). Clinical validation of a new test for aphasia: An experimental study on the Token Test. *Cortex, 2*(2), 222–243.

Romani, C., & Galluzzi, C. (2005). Effects of syllabic complexity in predicting accuracy of repetition and direction of errors in patients with articulatory and phonological difficulties. *Cognitive Neuropsychology, 22*(7), 817–850. doi: 10.1080/02643290442000365

Square, P. A., Roy, E. A., & Martin, R. E. (1997). Apraxia of speech: Another form of praxis disruption. In L. J. G. Rothi & K. M. Heilman (Eds.), *Apraxia: The neuropsychology of action* (pp. 173–206). Hove, UK: Psychology Press.

Staiger, A., & Ziegler, W. (2008). Syllable frequency and syllable structure in the spontaneous speech production of patients with apraxia of speech. *Aphasiology, 22*(11), 1201–1215. doi: 10.1080/02687030701820584

Strand, E. A., Duffy, J. R., Clark, H. M., & Josephs, K. (2014). The apraxia of speech rating scale: A tool for diagnosis and description of apraxia of speech. *Journal of Communication Disorders, 51*, 43–50. doi: 10.1016/j.jcomdis.2014.06.008

Van der Merwe, A. (2009). A theoretical framework for the characterization of pathological speech sensorimotor control. In M. McNeil (Ed.), *Clinical management of sensorimotor speech disorders*, (2nd ed., pp. 1-25). New York, NY Thieme.

Varley, R. & Whiteside, S. (2001). What is the underlying impairment in acquired apraxia of speech? *Aphasiology, 15*(1), 39–84. doi: 10.1080/02687040042000115

Wambaugh, J. L. (2006). Treatment guidelines for apraxia of speech: Lessons for future research. *Journal of Medical Speech Language Pathology, 14*(4), 317-321.

Wertz, R., LaPointe, L., & Rosenbek, J. (1984). *Apraxia of speech in adults. The disorder and its management*. New York, NY: Grune and Stratton.

West, C., Bowen, A., Hesketh, A., & Vail, A. (2008). Interventions for motor apraxia following stroke. *Cochrane Database of Systematic Reviews 2008, Issue 1*. doi: 10.1002/14651858.CD004132.pub2

World Health Organization. (2005). *The International Classification of Functioning, Disability and Health (ICF)*. Retrieved from http://www3.who.int/icf/.

Ziegler, W. (2002). Task-related factors in oral motor control: Speech and oral diadochokinesis in dysarthria and apraxia of speech. *Brain and Language, 80*(3), 556–575.

Ziegler, W. (2003). Speech motor control is task-specific: Evidence from dysarthria and apraxia of speech. *Aphasiology, 17*(1), 3–36. doi: 10.1080/729254892

Ziegler, W. (2005). A nonlinear model of word length effects in apraxia of speech. *Cognitive Neuropsychology, 22*(5), 603-623. doi: 10.1080/02643290442000211

Ziegler, W. (2008). Apraxia of speech. In M. J. Aminoff, F. Boller, D. F. Swaab, G. Goldenberg, & G. L. Miller (Eds.), *Handbook of clinical neurology: Neuropsychological and behavioral neurology* (Vol. 88, pp. 269–285). Amsterdam, The Netherlands: Elsevier.

Ziegler, W., Aichert, I., & Staiger, A. (2012). Apraxia of speech: Concepts and controversies. *Journal of Speech, Language, and Hearing Research, 55*(5), 1485–1501. doi: 10.1044/1092-4388(2012/12-0128)

Acknowledgements

The Dutch Aphasia Foundation (SAN) and the Beatrixoord Noord Nederland Foundation sponsored this study. We would like to thank the participants and the S-LPs that took part in this study. Furthermore, we would like to thank Dörte de Kok and Matt Coler for their comments on an earlier version of this article.

Declaration of interest

Two of the authors were involved in the development of the Diagnostic Instrument for Apraxia of Speech (DIAS; Feiken & Jonkers, 2012).

Authors' Note

Correspondence concerning this article should be addressed to Roel Jonkers, Department of Linguistics, University of Groningen, P.O. Box 716, 9700 AS Groningen, The Netherlands. Email: **r.jonkers@rug.nl**.

Appendix A
Items of the Subtests

Articulation of phonemes

	Consonants	Vowels
1.	/t/	/oo/
2.	/f/	/ee/
3.	/s/	/u/
4.	/h/	/a/
5.	/p/	/ie/
6.	/k/	/o/
7.	/r/	/uu/
8.	/l/	/i/
9.	/j/	/eu/
10.	/b/	/ei/
11.	/n/	/oe/
12.	/g/	/aa/
13.	/m/	/ui/
14.	/d/	/e/
15.	/w/	/ou/

Diadochokinesis

1. Pa pa pa

2. Pa ta ka

3. Mok mok mok

4. Mok sok hok

5. Dam dam dam

6. Dam das dak

7. Schel schel schel

8. Schel stel spel

9. Vlok vlok vlok

10. Vlok stok brok

11. Stank stank stank

12. Stank blank drank

Articulation of words

a. One syllable, not complex

1. sok (sock)

2. web (web)

3. kat (cat)

4. noot (nut)

5. veer (feather)

6. tas (bag)

b. Two syllables, not complex

1. kanon (canon)

2. minuut (minute)

3. banaan (banana)

4. debuut (début)

5. zadel (saddle)

6. gebak (cake)

c. One syllable, CC, 3 phonemes

1. knie (knee)

2. vlo (flea)

3. trui (sweater)

4. sla (salad)

5. prei (leek)

6. twee (two)

d. One syllable, CC, 4 phonemes

1. tand (tooth)

2. wesp (wasp)

3. punt (point)

4. gans (goose)

5. koord (cord)

6. bank (bank)

e. One syllable, CCC, 4 phonemes

1. arts (doctor)

2. sprei (bedspread)

3. angst (fear)

4. stro (straw)

5. oogst (harvest)

6. eerst (first)

f. One syllable, CCC, 5 phonemes

1. spraak (speech)

2. schrik (fright)

3. dorst (thirst)

4. schroef (screw)

5. kunst (art)

6. streep (line)

g. Two syllables, C-C, 5 phonemes

1. oksel (armpit)

2. pasta (pasta)

3. advies (advice)

4. omdat (because)

5. asbak (ashtray)

6. afweer (defense)

h. Three syllables, C-C, 8 phonemes

1. impulsief (impulsive)

2. abnormaal (abnormal)

3. aantasten (affect)

4. verwonden (wound)

5. onwaarheid (untruth)

6. inpalmen (to charm)

i. 4 syllables, not complex, 8 phonemes

1. televisie (television)

2. limonade (lemonade)

3. vitamine (vitamin)

4. politica (politician; fem.)

5. mayonnaise (mayonnaise)

6. apparatuur (apparatus)

j. not complex, 8\9-11 phonemes

1. fotocamera (photo camera)

2. kilometer (kilometre)

3. honorarium (fee)

4. figureren (figure; verb)

5. papegaaien (parrot; verb)

6. telefoneren (telephone; verb)

h. complex, 9-11 phonemes

1. invloedrijk (influential)

2. handtastelijk (palpable)

3. fietstassen (cycle-bags)

4. gras groeit (grass grows)

5. herfstblad (autumnal leaf)

6. eerstejaars (first-year student)

Appendix B
Determination of the Cut-off Scores

The *articulation of phonemes* subtest was used to assess two signs. To detect whether or not participants produce inconsistent realizations of phonemes, the number of inconsistencies within a three-time repetition was calculated (range: 0–30). To detect whether participants produced more errors in consonants than in vowels, the scores for correctly produced consonants and vowels were subtracted from each other (range: 0–15).

The *diadochokinesis* test was used to assess two signs. First, it was evaluated whether participants experience more difficulties in alternating than sequencing syllables and words, and secondly, it was observed whether participants show visible or auditory groping. For the assessment of the first symptom, the number of correct realizations in the repeating sequence (/pa-pa-pa/) was compared to those in the alternating sequence (/pa-ta-ka/). If the participants were able to perform the single repetition, they were asked to produce as many repetitions as possible in eight seconds. The number of correct realizations for the alternating sequence was then divided by the correct realizations for the repeated sequence, where a run of three syllables constituted a sequence. The obtained scores were increased by 1 in order to circumvent nil scores (range: unlimited). A score below 1 indicates a poorer performance on the alternating sequences.

The diadochokinesis test was also used to score the symptom of *groping*. This was done by scoring the presence of this symptom during the repetition of the alternating sequences.

The four remaining signs (initiation problems, syllable segmentation, segmentation of consonant combinations, and effect of articulatory complexity) were captured in the *articulation of words* subtest. Sixty-six words were divided into 11 blocks of increasing complexity. To prevent reliance on one single instance of a symptom, but also to keep scoring time within proportional limits, the presence of the signs was scored per block of six words. Initiation problems were scored in all blocks, so the highest score is the presence of 11 signs in 11 blocks (score 11/11 = 1). The other signs were only observed in a selected group of blocks. Syllable segmentation can only be observed in the polysyllabic words. There were 36 polysyllabic words, used in six blocks, so the highest score is 6/6 (score = 1). Segmentation of consonant clusters can only be observed in words including a consonant cluster. This was the case for 30 words, used in five blocks. The highest score is 5/5 (score = 1).

The effect of articulatory complexity was determined by comparing words of similar length but different articulatory complexity; two blocks contained non-complex words, two blocks contained a two-consonant cluster, and two blocks contained words with a three-consonant cluster. To account for an effect of articulatory complexity, the score of the third block was subtracted from the mean score of the three blocks.

KEY WORDS

LITERACY
SPELLING
ORAL LANGUAGE
SCREENING
SPEECH PERCEPTION
SPEECH PRODUCTION
PHONOLOGICAL AWARENESS
MORPHOLOGY

Development of a Tool to Screen Risk of Literacy Delays in French-Speaking Children: PHOPHLO

Développement d'un outil pour dépister le risque de retard dans l'acquisition des habiletés de littératie chez les enfants francophones : PHOPHLO

Susan Rvachew
Phaedra Royle
Laura M. Gonnerman
Brigitte Stanké
Alexandra Marquis
Alexandre Herbay

Susan Rvachew
School of Communication Sciences and Disorders, McGill University, Montréal, QC, CANADA
Centre for Research on Brain, Language and Music, Montréal, QC, CANADA

Phaedra Royle
École d'orthophonie et d'audiologie, Université de Montréal, Montréal, QC, CANADA
Centre for Research on Brain, Language and Music, Montréal, QC, CANADA

Laura Gonnerman
School of Communication Sciences and Disorders, McGill University, Montréal, QC, CANADA

Brigitte Stanké
École d'orthophonie et d'audiologie, Université de Montréal, Montréal, QC, CANADA

Alexandra Marquis
École d'orthophonie et d'audiologie, Université de Montréal, Montréal, QC, CANADA
United Arab Emirates University, Abu Dhabi, UNITED ARAB EMIRATES

Alexandre Herbay
School of Communication Sciences and Disorders, McGill University, Montréal, QC, CANADA

Abstract

Literacy is crucial for success, both professionally and personally. Oral language skills are closely related to literacy development in children. When a child has weak oral language skills, they will have difficulty achieving reading and writing competencies within the expected time frame. In this paper, we present results from a longitudinal and cross-sectional study of the relationship between oral language skills in pre-literate children, and one aspect of their literacy skills in early elementary school—specifically, spelling. The study was conducted with French-speaking children and French-language learners from Quebec, a population that has been understudied in this area. We developed a predictive tool that will allow teachers and other professionals to assess oral language skills in young children and to predict those children at risk for literacy difficulties. Specifically, we screened children's speech perception, speech production, phonological awareness, and morphology production abilities at entry to first grade and predicted spelling skills at the end of second grade. The screening tool that we developed proved to have a sensitivity of 71% and a specificity of 93% as a screen for poor spelling abilities.

Abrégé

La littératie est un élément crucial du succès à la fois professionnel et personnel. Les habiletés de langage oral sont intimement liées au développement de la littératie chez les enfants. En effet, lorsqu'un enfant a de faibles habiletés de langage oral, il aura plus de difficulté à développer ses habiletés de lecture et d'écriture dans les délais prévus. Nous présentons les résultats d'une étude longitudinale et transversale qui explore les liens entre les habiletés de langage oral chez des enfants n'ayant pas appris à lire ou à écrire et leurs habiletés de littératie au premier cycle du primaire. Cette étude a été menée auprès d'enfants franco-québécois natifs et non natifs, une population peu étudiée dans ce domaine. Nous avons créé un outil prédictif qui permettra aux enseignants et autres professionnels d'évaluer les habiletés de langage oral des enfants et de prédire ceux qui sont à risque de présenter des difficultés de littératie. Plus spécifiquement, nous avons évalué les habiletés de perception et de production de la parole, de conscience phonologique et de production morphologique d'enfants débutant leur première année du primaire. Nous avons prédit leurs habiletés d'orthographe à la fin de leur deuxième année (fin du premier cycle du primaire). L'outil développé a démontré une sensibilité de 71% et une spécificité de 93% pour dépister les faibles habiletés d'orthographe.

Literacy skills are essential for success in modern life, at the level of the individual and at the level of broader society. Stronger literacy skills are associated with a greater likelihood of school completion (Hernandez, 2011); furthermore, individuals with higher literacy skills have an employment and earnings advantage even after controlling for educational attainment (Organization for Economic Cooperation and Development [OECD], 2011). Literacy is also an important social determinant of mental and physical health (Dewalt, Berkman, Sheridan, Lohr, & Pignone, 2004; Marcus, 2006). At the society level, communities with a greater proportion of highly literate individuals enjoy a greater quality of life, not only in economic terms but also through enhanced social cohesion, as literacy is associated with greater civic participation (OECD, 2011). Speech-language pathologists (S-LPs) play a pivotal role in ensuring these positive outcomes for individuals and society, because oral language skills are the foundation of literacy and because S-LPs are key members of the team of professionals responsible for literacy outcomes in the school environment (Justice, 2006; Lefebvre, Trudeau, & Sutton, 2008; Roth & Baden, 2001). In this paper, we highlight the close relation between oral and written language skills and introduce a new screening tool to identify French-speaking children who are at risk for literacy delays without additional support. In Quebec, high dropout rates from secondary school are recognized as "a major problem" that is correlated with poor written language performance (Fortin, Royer, Potvin, Marcotte, & Yergeau, 2004). Screening, assessment and intervention tools that are adapted for the particular needs of the Canadian French-speaking population are urgently needed.

Written language skills in Quebec school children

With respect to literacy, French-speaking Canadian children tend to underperform compared to their English-speaking counterparts across Canada, as revealed by the *Programme international de recherche en lecture scolaire* (PIRLS; Labrecque, Chuy, Brochu, & Houme, 2012), which tracks fourth-grade reading competence on a regular basis, permitting comparisons across language groups, genders, and provinces. PIRLS results from 2011 show that Canada as a whole and Quebec as a province score significantly higher than the world average. However, Quebec students from French-language school boards underperform compared to the Canadian average and the average of students in English-language schools in Quebec. (As an aside, students in minority French-language school boards elsewhere in Canada underperform compared to the Canadian French-language average). The state of literacy in Quebec has been a major concern for some years now, since a government report revealed significant difficulties in children's writing abilities in primary and secondary school (Gouvernement du Québec, 2006). Even more worrying, students' writing skills at the end of sixth grade were statistically weaker in 2005 than those of their peers five years earlier (Jalbert, 2007). Subsequently, the Ministère de l'éducation [Ministry of Education] introduced a new approach to literacy education in Quebec that included a competency-based approach to the teaching and assessment of reading and writing.

The literacy skills of Quebec school children are assessed through obligatory province-wide writing assessments administered in primary school (fourth and sixth grade), with additional tests in secondary school. Over several days, the students read and discuss a variety of texts and then write a narrative (in primary school) or explanatory text (in secondary school). These written texts are graded for relevance, organization, syntax and punctuation, vocabulary, and orthography. Each of these five areas is rated separately as very satisfactory, satisfactory, acceptable, somewhat satisfactory, or unsatisfactory, according to specific criteria. For example, syntax and punctuation is scored globally, so that an "unsatisfactory" rating indicates that sentence structure and punctuation rarely met expectations throughout the text. However, orthography is scored by counting the exact number of errors on a word-by-word basis, taking into account spelling and grammatical errors at the word level; scoring grids are provided by grade and text length, such that a "very satisfactory" fourth-grade text would contain less than 4% incorrect words. Provincial reports focus on rate of success (percentage of students receiving at least "acceptable" ratings), as well as percentage of "unsatisfactory" ratings, which indicate the need for special resources in the system.

A report on recent student performance on the obligatory writing tests from June 2009 (Charest, 2010) revealed that boys scored significantly lower than girls on average and across all scoring criteria. The rate of success declined with age (from 81% to 68%). The decline with age was particularly marked for orthography; furthermore, for both younger and older children, the disparity between boys and girls was most noticeable in this area. In the primary grades, the lowest rate of success was for syntax and punctuation but "unsatisfactory" ratings occurred most often for orthography. The distribution of scores in the orthography category was noticeably bimodal, with many children achieving "very satisfactory" scores but a substantial group showing "unsatisfactory" performance in this area (grade school: 3% of girls and 9% of boys; high school: 5% of girls and 10% boys). Motivation to read was

a significant predictor of outcomes for younger children, and perceived competence in reading and writing were significant predictors of outcomes for older children.

In this report we will focus on spelling as the literacy skill of interest given that it emerged as a particular area of difficulty on the obligatory literacy assessment in Quebec (Charest, 2010). Furthermore, it is an early marker of more generalized difficulties with writing and literacy overall. Spelling may be a particularly sensitive indicator of literacy problems; several studies have shown that at-risk children who have poor reading skills are usually poor at spelling, whereas some children are poor spellers while having relatively good reading skills (Holm, Farrier, & Dodd, 2008; Lewis, Freebairn, & Taylor, 2000; Pennala et al., 2010). Furthermore, follow-up of participants being treated for dyslexia indicates the persistence of spelling and writing difficulties long after resolution of the reading impairment, in children (Berninger, Nielsen, Abbott, Wijsman, & Raskind, 2008) and in adults (Connelly, Campbell, MacLean, & Barnes, 2006). Finally, some studies have demonstrated a positive impact of spelling instruction on reading and other literacy skills (Graham & Santangelo, 2014; Weiser & Mathes, 2011).

Oral language foundations of literacy

Literacy includes a host of interconnected skills involving print: letter and letter-sound knowledge, decoding and sight word reading, spelling, grammatically correct and coherent writing of sentences and passages, reading fluency, reading comprehension, and ultimately the ability to gain new knowledge and solve problems using print materials, whether in paper or digital form. These written language skills are learned through direct teaching and practice, beginning in preschool but with particularly explicit attention devoted to the teaching of reading and writing during the early school years. However, the foundation for literacy is formed during the preschool period with the acquisition of oral language skills, beginning with language-specific shaping of perceptual knowledge during the first year of life. Every aspect of literacy has been shown to be closely correlated with oral language skills, including decoding, reading fluency, and reading comprehension (Durand, Loe, Yeatman, & Feldman, 2013). Furthermore, children who have speech and language impairments are at risk for delayed acquisition of literacy (Puranik, Petcher, Al Otaiba, Catts, & Lonigan, 2008).

Longitudinal studies have linked oral language development during the preschool period to the acquisition of literacy skills after school entry (Cooper, Roth, Speece, & Schatschneider, 2002; Dickinson, McCabe, Anastasopoulos, Peisner-Feinberg, & Poe, 2003; Dickinson & Porsche, 2011; Hulme, Nash, Gooch, Lervag, & Snowling, 2015; Sénéchal & LeFevre, 2002; Snowling & Melby-Lervåg, 2016; Speece, Roth, Cooper, & de la Paz, 1999; Storch & Whitehurst, 2002). These studies show that oral language skills exert both indirect and direct effects on literacy. First, phonological awareness emerges from accumulating knowledge in the phonological and lexical domains; in turn, phonological awareness (which emerges implicitly) and letter-sound knowledge (which must be taught explicitly) combine to underpin the child's acquisition of decoding skills. In this way, oral language skills exert an early indirect effect on the earliest stages of literacy acquisition. Later, when the child is "reading to learn", oral language abilities—such as vocabulary, syntax, and oral narrative abilities—directly support written language comprehension (Griffin, Hemphill, Camp, & Palmer Wolf, 2004; Nation & Snowling, 2004). An indirect effect remains because speed and automaticity in the decoding process support comprehension when reading sentence and passage level text. Oral language skills also support the child's writing abilities at every level, including spelling, syntax, and narrative structure (Stothard, Snowling, Bishop, Chipchase, & Kaplan, 1998).

Prediction of spelling abilities

Given the heightened and continuing concern about the written language skills of French-speaking children in Quebec, a targeted funding program was implemented to encourage research in this area. Consequently, we embarked on a project to develop a screening tool that could be used to identify children at school entry who would potentially be at risk for slower acquisition of writing, or more specifically in this context, spelling at the end of second grade. For predictors, we chose four aspects of oral language abilities that are known to be correlated with spelling specifically and literacy more generally. These predictors are discussed in turn below: speech perception, speech production, phonological awareness, and morphology production.

Speech perception skills are a known correlate of emergent literacy skills, reading ability, and spelling (Anthony, Lonigan, Driscoll, Philips, & Burgess, 2003; Boets, Wouters, van Wieringen, De Smedt, & Ghesquière, 2008; Overby & Bernthal, 2008). For example, the ability to discriminate short versus long vowels was found to be associated with literacy skills in Finnish children, with second-grade spelling abilities being the strongest correlate (Pennala et al., 2010). A speech perception test that uses a word identification procedure appropriate for young

children, and which was previously validated as a predictor of phonological awareness and emergent literacy skills in English—the Speech Assessment and Interactive Learning System (SAILS)—was modified for the French context and used in this study (Rvachew & Grawburg, 2006).

Speech production accuracy also influences children's spelling abilities. For example, articulatory similarity of vowels explains spelling confusions in English and in French (Caravolas & Bruck, 2000; Ehri, Wilce, & Taylor, 1987). Furthermore, underlying organization of phonological structure also explains common error patterns in early spelling (Bourassa & Treiman, 2001), such as, for example, the omission of word-internal nasals or liquids that are represented as vocalic rather than consonantal elements (e.g., "hand" ⟶ "had"). Many studies have shown that children who present with a speech sound disorder are at risk for future difficulties with spelling, even when their language abilities are within the average range (Bird, Bishop, & Freeman, 1995; Lewis, Freebairn, & Taylor, 2002; Overby, Masterson, & Preston, 2015). Therefore, we included a test of speech production accuracy that has been used to describe the speech abilities of monolingual and bilingual children in primary school (the *Test de Dépistage Francophone de Phonologie* [TDFP]; Rvachew et al., 2013), as well as the speech errors produced by preschoolers with a phonological disorder (Brosseau-Lapré & Rvachew, 2014; Paul, 2009). Moreover, performance on this test has been shown to be closely related to phonological awareness performance (Brosseau-Lapré & Rvachew, 2017).

Phonological awareness is well recognized as an excellent predictor of reading and spelling abilities (Holm et al., 2008; Schneider, Roth, & Ennemoser, 2000). For example, Speece et al. (1999) found that strong phonological skills in kindergarten were associated with strong spelling abilities in first grade. We selected a measure of implicit phonological awareness skills, requiring no spoken responses, so that the children's performance would be independent of their speech accuracy. The English version of this test predicts reading and spelling ability (Bird et al., 1995; Rvachew, 2007). The French version—the *Test de Conscience Phonologique Préscolaire* (TCPP)—has previously been used to describe and differentiate phonological awareness skills of children receiving speech therapy from children with normally developing speech and language skills (Rvachew & Brosseau-Lapré, 2015; Brosseau-Lapré & Rvachew, 2017).

The fourth target of our screening protocol was knowledge of grammatical morphemes, specifically the past tense. As previously mentioned, aspects of grammar were particularly difficult for French-speaking children on their writing tests. Furthermore, expressive morphology is an aspect of structural language development that has been previously linked to literacy development in general (Speece et al., 1999). Morphological awareness emerges in early primary grades to aide reading and writing (Duncan, Colé, & Casalis, 2009; Pacton & Deacon, 2008; Wolter, Wood, & D'zatko, 2009). For example, Sénéchal, Basque, and Leclaire (2006) showed that morphological awareness was correlated with the ability to spell morphological and lexical words in grade 4 French-speaking children. Metalinguistic knowledge of inflectional and derivational morphology is particularly helpful to spelling, but explicit morphological awareness is more reliably assessed in second and third grade compared to early first grade (Bédard, Marquis, Royle, Gonnerman, & Rvachew, 2013). Therefore, we included a measure of productive morpheme knowledge that we have used previously to describe the development of morphology in young Quebec children with and without language impairments (*Jeu de Verbes*; Marquis, Royle, Gonnerman, & Rvachew, 2012; Royle, 2007; Royle & Thordardottir, 2008). This test assesses the child's ability to produce French verbs in the *passé composé* (perfect past) form, using the auxiliary *avoir* ("to have") or *être* ("to be") and a past participle of the verb. We used this specific structure because it is acquired early (Thordardottir & Namazi, 2007) and it can be reliably elicited in children as young as age 3;2 (years;months; Royle, 2007). In contrast, many aspects of morphology are highly irregular or are variably produced in oral French (Kresh, 2008; Legendre et al., 2009). Other aspects of morphology that involve allomorphy (e.g., liaison, elision, and contraction; Béchara, 2015) were not tested because they confound morphological and phonological processes.

We chose spelling at the end of second grade as our outcome, given that spelling is an area of particular weakness, and spelling may be an early indicator of the writing difficulties identified throughout the school years on the province-wide literacy competency assessment. Therefore, word and phrase level spelling was tested from dictation, using the *Batterie d'évaluation de lecture et d'orthographe* (BELO; George & Pech-Georgel, 2006), as the final outcome at the end of second grade. The BELO was standardized on a sample of 371 early-grade children and found to have excellent reliability and convergent validity. In particular, the BELO was validated against the Alouette (Lefavrais, 2006) on 100 children (Pech-Georgel & George, 2010). This task was chosen because it is adapted to the age level and language of our participants and evaluates phono-orthographic abilities (non-word syllables), basic orthographic abilities for known words

(real words), and basic grammatical abilities (sentences). Although the test is based on dictation rather than free narrative (as is developmentally appropriate for second-grade spellers; see Alamargot, 2007), the coding is similar to that used in the provincial writing assessment, in that each word is scored as spelled correctly or incorrectly, capturing spelling and grammatical abilities simultaneously.

Overview and objectives

The screening test was developed in a two-phase process. This research project will be described in relation to the objectives for each of the two phases, as follows:

Phase I, Objective 1: Administer the full battery of assessments to kindergarten and first-grade children in order to test whether our measures of speech perception, speech production, phonological awareness and morphology production would differentiate children likely to differ in writing abilities as consequence of variations in grade, language background, perceived risk, and overall test performance.

Phase I, Objective 2: Using item-level discriminability and difficulty statistics, select a smaller set of items from among these measures to form a screening test, which is hypothesized to predict future spelling abilities while being shorter than the full test battery.

Phase II, Objective 3: Administer the screening test, *Prédiction des Habiletés Orthographiques Par des Habiletés Langage Oral* (PHOPHLO), to first-grade children, followed by a spelling test, BELO, to the same children in second grade, in order to determine the specificity and sensitivity of the PHOPHLO as a predictor of BELO performance.

Phase II, Objective 4: Examine the contribution of each of the four subtests in the screener to the identification of children who proved to have poor spelling performance at the end of second grade, with the expectation that the test as a whole and the individual subtests will contribute to the prediction of spelling abilities.

General Method

Testing protocols were approved by the internal review boards from both the Université de Montréal and McGill University Faculties of Medicine. The children were recruited from their school by sending letters home and asking parents to return a signed consent form if they agreed to their child's participation.

The study participants were drawn from a French public school board located in a suburb of Montréal in the province of Quebec (Canada). The particular area from which the children were recruited, according to the most recent census, is an area of high immigration with 61% of the total population speaking French as the mother tongue and 28% speaking neither English nor French as the mother tongue. Less than 13% of the population speaks English regularly at home. By law, immigrant children must be educated in French in Quebec.

All children in the kindergarten and first-grade classrooms were eligible for participation regardless of their language background or the presence of developmental difficulties, as long as the parent consented and the child assented and was able to cooperate with the testing procedures. A telephone interview was conducted with each child's parent to obtain demographic, literacy, health, and language information via standard questionnaires. Parents identified possible developmental concerns for some children but we did not verify these concerns via diagnostic testing or by obtaining confirmatory documentation. Language status was based on parental reports of their own language use with and around their child, siblings' language use, other caregivers' language use, and radio and television exposure. A 90% criterion of French exposure from birth was used to determine monolingual status of children placed in the monolingual (ML) group. The remaining children were placed in the bilingual (BL) group (i.e., either simultaneous BL with exposure to two languages from birth, or sequential BL with no French exposure until preschool). The languages represented besides French were diverse, including English, Arabic, Spanish, Haitian Creole, Italian, Greek, Lao, Polish, Romanian, Asu, and Khmer. Teachers were also asked to rate each child as being "at-risk" or "not-at-risk" for developing writing difficulties, on the basis of their own opinion with no specific criteria provided (for more information about the teacher ratings, see Kolne, Gonnerman, Marquis, Royle, & Rvachew, 2016).

Children were tested individually in a quiet room inside the school. The assessment protocol in both phases was administered by native French-speaking graduate level research assistants under the supervision of a post-doctoral fellow, the fifth author. All scoring, transcription, and reliability coding was subsequently completed by native French-speaking graduate students in speech-language pathology with training in clinical phonetics and phonology, under the supervision of the first and second authors.

Phase I: Development of screening test

Method

The Phase I experiment involved cross-sectional

assessments of children in kindergarten and first-grade classrooms. The children were tested at the end of the school year. Although the final screening test is intended to identify children who may be at risk for written language problems prior to onset of formal reading instruction, a group of children who were expected to have beginning reading skills were included in the Phase I sample (that is, children at the end of the first-grade year). This was so that the items for the screener could be selected that discriminated performance across a broad range of skill levels.

Participants

The children recruited to the Phase I experiment comprised 43 children from kindergarten classrooms with a mean age of 6 years and 1 month, including 21 boys (22 girls) and 24 ML (19 BL) speakers of French. From a first-grade classroom, 18 children were recruited with a mean age of 7;2, including 11 boys (seven girls) and 12 ML (six BL) speakers of French. On average, the number of years of maternal education was 14.48 (*SD* = 2.06). Developmental diagnoses were suspected but not confirmed by professional assessments for four kindergarten children (autism spectrum disorder, intellectual disability, and attention deficit disorder). A heart defect was reportedly diagnosed for one child and language impairment for another. Parents reported concerns about hearing due to otitis media for five children and about fine motor skills for three children.

Procedure

The four different language assessment tasks were administered over two separate sessions pairing phonological awareness and speech perception in one 20-minute session, and morphology production and phonological production in another 20-minute session, with order counterbalanced within session and across participants. Evaluation sessions occurred within a maximum of two weeks from each other. Sessions were recorded with a Zoom1 stereo digital recorder at a sampling frequency of 44 kHz and a quantization rate of 24 bits. Responses to the speech perception and phonological awareness tasks were automatically recorded by the test device, whereas responses to the speech production and morphology production tasks were transcribed from audio recordings. Subsequently, 16% of all audio recordings for these two production tasks were retranscribed to obtain estimates of transcription reliability. Following data collection and coding, the children's performance on each test item was examined to reveal item difficulty and item discrimination scores when differentiating risk status (according to the teacher rating) and grade (kindergarten versus grade 1) and overall performance (using a split-half procedure for total test score regardless of child's age or grade or risk status).

Speech perception. The Speech Assessment and Interactive Learning System (SAILS; Rvachew, 2009) assesses speech perception with a two-alternative, forced-choice word identification task. The child hears natural speech recorded from adults and typically developing children. The words are presented in blocks of 10 items, five representing the target and five representing a misarticulated version of the target word. The child listens to each word and points to a picture of the target when a correct pronunciation is heard and an X when a misarticulation is heard. A laptop was used to run the software that ensures random ordering of stimuli within blocks. The child listened to the stimuli over headphones, presented at the loudest comfortable level. The examiner used a mouse to activate the hotspot selected by the child on the computer screen and responses were recorded automatically by the software. A reinforcement image was presented after each response, regardless of whether the child's response was correct or not. An experimental French version of SAILS was developed for this study, which included two blocks of *gris* ([gʁi] – "grey") stimuli recorded from preschool-aged children, two blocks of *serpent* ([sɛʁpã] – "snake") stimuli recorded from adults, and two blocks of *poisson* ([pwasɔ̃] – "fish") stimuli recorded from adults. Erroneous tokens represented commonly occurring misarticulations, including omissions (e.g., *gris* → [gi], *poisson* → [pasɔ̃], *serpent* → [sɛpã]) and substitutions of consonants (e.g., *gris* → [gji], *poisson* → [bwasɔ̃]) and vowels (e.g., *serpent* → [sɛʁpɑ]). Each test was preceded by 10 practice trials involving an easy contrast (e.g., [gʁi] versus [mi]), during which the examiner could help the child to understand the task. The test comprises 60 items, of which 30 are practice items. The test is scored as percentage of items correct out of the remaining 30.

Speech production. The *Test de Dépistage Francophone de Phonologie* (TDFP), described in complete detail in Rvachew et al. (2013), comprises eight colour photos, presented digitally with verbal prompts used to elicit 30 spoken words from the child. The words were selected to be known by children aged 2 to 8 and to be representative of the distribution of phonemes, syllable shapes, and word lengths characteristic of Quebec French. Consonants appear in four syllable positions: singleton syllable onset (e.g., the first consonant in *niche* [niʃ] – "doghouse"), branching onsets (e.g., the two consonants at the beginning of the word *clown* [klun] – "clown"), glide in the nucleus (e.g.,

glide following the /v/ in *avion* [avjõ] – "plane") and the coda (e.g., the /ʁ/ in the words *serpent* [sɛʁpã] and *hélicoptère* [elikɔptɛʁ] – "helicopter"). The test is scored as percentage of consonants correct with every consonant in every word considered, therefore 94 consonants within 30 words. Inter-rater point-by-point transcription agreement for narrow transcription of consonants was 95.9%.

Phonological awareness. The *Test de Conscience Phonologique Préscolaire* (TCPP; Brosseau-Lapré & Rvachew, 2008) was modeled on the phonological awareness test developed by Bird et al. (1995), adapting it for French and implementing it on a computer using HTML software code. It consists of three subtests but only the first (rime matching) and third subtests (onset segmentation and matching) were administered to the children in this study. (The second subtest also targets onset matching and was omitted to reduce testing time because this is a very long assessment). In the rime matching subtest, the child is presented with an animal and its name, and told that it "likes things that sound like his name". In the onset and segmentation subtest the child is told the animal "likes things that start with the same sound as his name". For each trial the child is presented with four pictured items (the target and three distractors) and these items are named for the child on every trial. There are five practice items at the beginning of each subtest during which corrective feedback can be provided as necessary. The task was presented by computer although the examiner provided extra support, especially during the practice items. The child responded by touching the appropriate picture and the software recorded responses automatically. The total test score is the number of correct items out of 24 (14 rime matching and 10 onset segmentation, excluding practice trials).

Morphology production. The children's ability to produce *passé composé* forms was assessed with an elicited production task for verbs using an interactive Android platform. The application simulated a storybook where the children are asked to complete short stories by responding to questions from the experimenter. The adults would read three short sentences presenting the target verb in order to induce the perfect past. For example, along with an image of a girl hiding her dolls under a box, the script presented was: *Marie va cacher ses poupées. Marie cache toujours ses poupées. Qu'est-ce qu'elle a fait hier Marie?* ("Marie will hide (infinitive) her dolls. Marie hides (present, 3rd person singular) her dolls every day. What did she do yesterday, Marie?") The tasks had four types of verbs with seven items each (four of which were used as practice items): seven verbs with a past participle in *–é* (/e/; e.g., *caché* – "hidden"); seven with a participle in *–i* (/i/; e.g., *fini* – "finished"); seven with a participle *–u* (/y/; e.g., *mordu* – "bitten"); and seven with other non-paradigmatic, or opaque, forms (e.g., *ouvert* – "opened"). All items are conjugated with *avoir*. The expected pronoun is *il* ("he") or *elle* ("her"), but was not counted as incorrect if a gender error occurred. The items are described in more detail in Marquis et al. (2012). One point was given for each correct production of the full *passé composé* (i.e., the pronoun clitic, auxiliary, and past participle; for example, (*Marie*), *elle a caché* – (Marie), she AUX hid.pp). The total score was out of 24. Coding reliability for correct production of 25% of tested children was 98.8%.

Results and Discussion

The children's performance, on average, for the four oral language tests, is shown in Table 1 by subtest for the full group and for contrasting subgroups, specifically kindergarten versus first-grade children, boys versus girls, BL vs. ML children, at-risk versus not-at-risk children according to teacher report, and low-scoring versus high-scoring children. The low- versus high-scoring subgroups were identified by transforming the scores on all four tests to *z*-scores, taking the mean of the *z*-scores across the four tests, and then splitting the whole group (kindergarten and first grade combined) at the median *z*-score. Differences in means across pairs of subgroups were assessed against the standard deviation of subtest scores for the full group of children. If one considers a half-standard deviation difference in means to be of interest, Table 1 shows that the tests were generally discriminating. Specifically, the speech perception test (SAILS) differentiated sub-groups on the basis of grade and overall test score (i.e., low vs. high scores). Speech production accuracy (TDFP) also differentiated kindergarten from first-grade children and the low-scoring from high-scoring subgroups. Phonological awareness (TCPP) differentiated groups well with differences between mean scores sometimes more than a standard deviation apart and differences apparent between grades, risk subgroups and low- versus high-scoring subgroups. Morphology production (*Jeu de verbes*) differentiated the BL versus ML subgroups.

Given that each of the four domains assessed proved to have some value for differentiating subgroups of children who might be expected to present with varying oral language skills, it was decided to include all four subtests after item analyses to reduce the length of testing. Detailed item analyses, including discriminability and difficulty indexes and item-total correlations, were used to select a smaller subset of items from each subtest to create a pilot screening tool that could be administered in a

Table 1. Results Obtained for the Four Oral Language Tests Administered in Phase I by Subgroup

		Speech Production		Speech Perception		Phonological Awareness		Morphology Production	
Subgroup	*N*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
All children	61	87.32	8.24	90.69	6.84	17.11	5.24	13.61	7.27
Kindergarten, vs.	42	**85.97**	8.94	**89.60**	7.48	**15.23**	5.01	12.74	7.42
First grade	18	**90.56**	5.14	**93.28**	4.07	**21.61**	2.15	15.67	6.63
Boys, vs.	29	86.88	6.81	90.31	6.47	17.16	5.23	14.59	7.01
Girls	32	87.82	9.68	91.10	7.31	17.07	5.25	12.52	7.52
L2 French, vs.	36	85.73	9.74	88.96	7.91	15.92	5.51	**11.52**	7.41
L1 French	25	88.43	6.95	91.89	5.80	17.94	4.96	**15.16**	6.90
At-risk, vs.	29	85.40	9.78	89.17	6.18	**14.48**	5.11	14.03	7.44
Not-at-risk	32	89.06	6.21	92.06	7.30	**19.50**	4.16	13.22	7.21
Low score, vs.	23	**82.32**	4.79	**88.35**	7.53	**15.43**	5.14	13.26	7.60
High Score	38	**90.35**	8.46	**92.11**	6.05	**18.13**	5.11	13.82	7.16

Note. Speech Perception is scored as percent correct over 30 items; Speech Production is scored as percent correct over 30 words and 94 consonants; Phonological Awareness is scored as number correct over 24 items; Morphology Production is scored as number correct over 24 items. Bold lettering highlights subgroup means that differ by more than one-half standard deviation (calculated from all children by test).

single test session. Specifically, items with relatively poor discriminability and difficulty indexes were eliminated from each subtest, using the top and bottom quarter of the sample, based on total subtest scores (Burton, 2001; Kelley, 1939; Gelman & Park, 2001). For example, with respect to the speech perception test, many of the *poisson* items were identified correctly by all of the children and therefore did not discriminate high- and low-scoring subgroups; in contrast, the five items that were ultimately selected for identification as incorrect exemplars of the word *gris* were associated with an average discrimination index of 0.37. Similarly, when considering the morphology production test, the verb item *couvrir* ("to cover") was eliminated with a discrimination index of 0.15, reflecting the extreme difficulty of this item for both high- and low-scoring children; by contrast, the item *remplir* ("to fill") was retained with a discrimination index of 0.85. This process was applied to each item in all four subtests, with the result described below by subtest.

The game *Écoute* ("Listen") tests speech perception with a 10-item word recognition procedure targeting the word *gris* in which five items are correctly produced ([gʁi]) and five items are misarticulations as follows: [gi], [gi], [gi], [gji], [ŋgi]. All items are produced by different child talkers so that even though some items are phonetically the same, each item is acoustically distinct. A practice block of 10 trials precedes the test block. A screenshot of a single practice trial is shown in Figure 1. During test trials, caterpillars turn into butterflies with each completed item, providing noncontingent feedback that helps the child gauge progress toward game completion.

The game *Qu'est-ce que c'est?* ("What is it?") tests speech production accuracy by presenting children with colour drawings of 10 items for naming. There are no practice trials but additional verbal prompts are available if the child does not know the name of the item. Specifically, if the child produces no response or the wrong word, the first

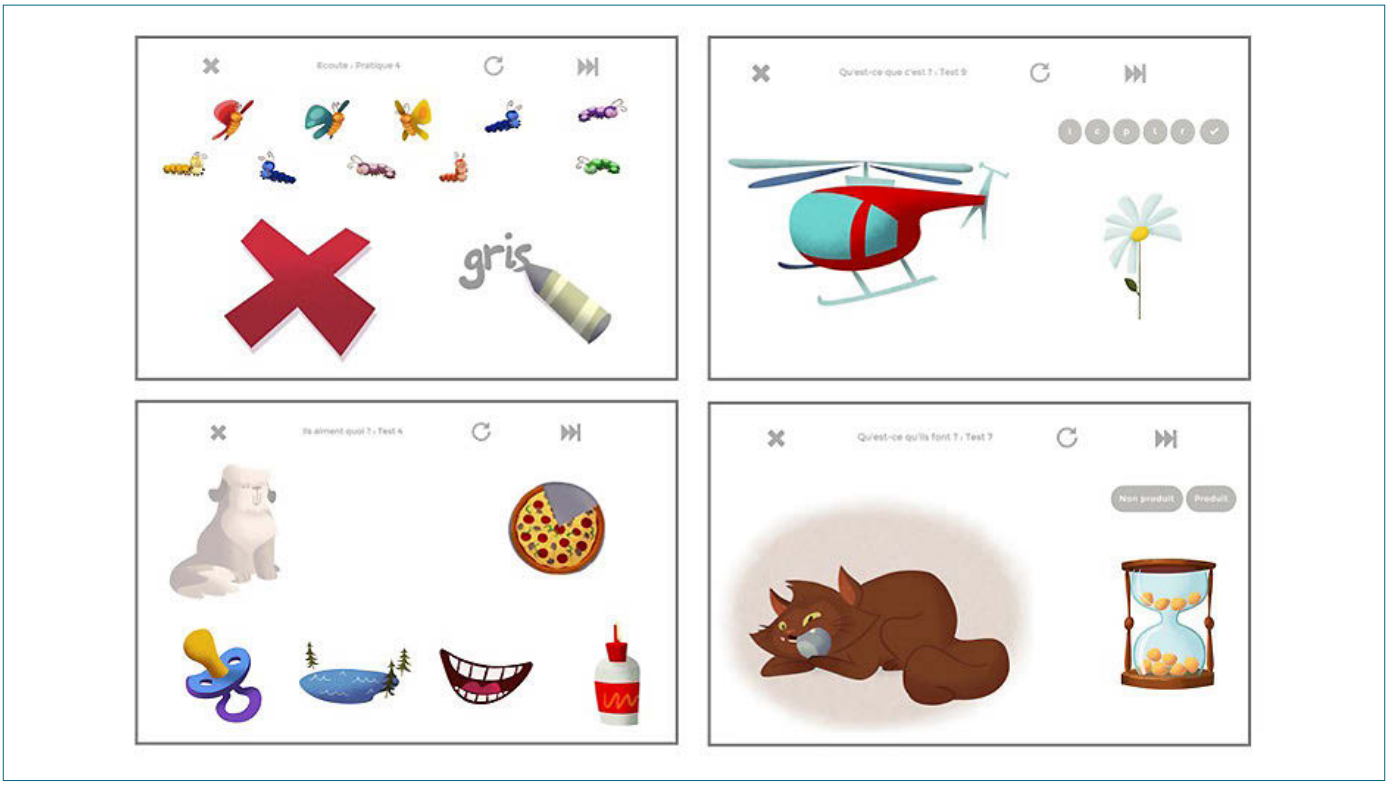

Figure 1. Screen shots from the four PHOPHLO subtests: Speech Perception, *Écoute* "Listen" (top left); Speech Production, *Qu'est-ce que c'est?* "What is it?" (top right); Phonological Awareness, *Ils aiment quoi?* "What do they like?" (bottom left); and Morphology Production, *Qu'est-ce qu'ils font?* "What are they doing?" (bottom right).

prompt provides a semantic hint and the second hint prompts for delayed imitation of the target word. The consonants in the word are presented on screen so that the examiner can then identify consonants that were misarticulated by the child, or, alternatively, the entire word can be marked as correct. The software provides a response grid to the examiner for recording production errors, yielding a count of correctly produced consonants out of 36 in total. The 10 items are *niche, tournevis* ("screwdriver"), *serpent, clown, araignée* ("spider"), *enveloppe* ("envelope"), *garde-robe* ("closet"), *parapluie* ("umbrella"), *hélicoptère*, and *camion* ("truck"). A screenshot of the item *hélicoptère* is shown in Figure 1. Daisy petals are added with each item to help the child gauge progress toward completion of the game.

The game *Ils aiment quoi?* ("What do they like?") tests rime awareness using the procedure previously described in which the child identifies the item that matches the rime of the name of the animal. Five practice trials using the names *Guy* ([gi]) and *Jeanne* ([ʒan]) are provided for teaching the task. Subsequently, 14 test items target the names *Lou* ([lu]), *Paul* ([pɔl]), *Lucas* ([luka]), and *Plé* ([plɛ]), in each case with four pictures shown representing the answer and three distractors. The software records the child's picture touch responses and sums correct responses for the test items. A screenshot showing the layout from one of the test items is shown in Figure 1. Disappearing pizza slices mark progress toward the end of the game, indicating trial completion without regard for response accuracy.

The game *Qu'est-ce qu'ils font?* ("What are they doing?") prompts production of *passé composé* verb forms using the procedure previously described. Ten items target the verbs *rire* ("to laugh"), *sentir* ("to smell"), *remplir, ouvrir* ("to open"), *conduire* ("to drive"), *battre* ("to beat/win"), *défendre* ("to defend"), *perdre* ("to lose"), *mordre* ("to bite"), and *boire* ("to drink"), most ending in *–i* or *–u* and one having an idiosyncratic form. Tablet icons permit the examiner to indicate which parts of the child's response were correct (subject + auxiliary + participle). The software provides detailed information about the child's performance (i.e., subject, auxiliary, and participle for each item), but the total score tabulated by the software reflects the number of complete items produced correctly, out of 10. Scoring thus reflects the child's ability to produce not only the past participle morpheme but to produce it in context including the subject and auxiliary. Again, feedback marking trial completion is noncontingent except for the practice trials. One trial from this game is shown in Figure 1 (specifically the trial that elicits *Il a mordu* – "He bit").

This process of item selection by subtest resulted in a 64-item screener (44 test trials and 20 practice trials) that correlated with the full test battery, $r = .89$, $p < .0001$. Test order is fixed but item order may be randomized within each test. The screener was subsequently developed as an integrated software tool that can be accessed and implemented on multiple digital platforms (Android or iPad tablets, or Windows or Macintosh computers). All visual and auditory stimuli are presented by the software, noncontingent visual feedback is provided for every child response, and these are recorded and tabulated by or with the assistance of the software. After the screening is complete, a complete record of the child's responses is provided along with an indication of whether the child passed or failed according to the criteria developed in Phase II of the project.

Phase II: Testing of an oral language screen as a longitudinal predictor of spelling

Method

The Phase II experiment involved longitudinal assessments of children tested during the first term of grade 1, using the PHOPHLO screener that was developed in Phase I, and again during the final term of grade 2, using the BELO test of spelling.

Participants

The children recruited to the Phase II experiment comprised 91 children from first-grade classrooms with a mean age of 6;9, including 36 boys (55 girls) and 52 ML (39 BL) speakers of French, with mean maternal education 14.55 years (*SD* = 2.08). Concerns about the children's development were raised in several areas, specifically language learning (three children), hearing (two children), attention deficits (four children), fine motor skills (three children), dyslexia (three children), social problems with peers (one child) and anxiety (one child). At the end of the second grade, 78 children were located to receive the outcome assessment. Some children were lost to follow-up because they moved out of the school district. Some children were included in a pilot study that involved first-grade administration of the BELO (Kolne et al., 2016), and therefore these children were excluded from the Phase II experiment. The 78 remaining participants were aged 8;2 on average, with the group composed of 30 boys (48 girls) and 45 ML (33 BL) speakers of French, and mean maternal education 14.46 years (*SD* = 2.09).

Procedure

At the beginning of first grade, the children were assessed with the PHOPHLO screening test, described as the outcome of the Phase I study, in a single session lasting approximately 20 minutes. As the children were approaching the end of second grade, the BELO (George & Pech-Georgel, 2006) was administered to the children in small groups of three or four, in order to assess their spelling ability. A standard dictation procedure was used: the examiner presented the items live-voice; the children wrote down what they heard on paper marked with familiar primary school line markings. The test was not timed and therefore each item was presented when the entire group had completed their transcription of the previous item. In the first section there were 10 non-word items: five single-syllable items (e.g., *fir*) and five two-syllable items (e.g., *palon*). Next, 15 real word items were presented, including 10 high-frequency words and five low-frequency words with simple (e.g., *fam* in *famille* – "family"), complex (e.g., *ille* in *famille*), and contextual grapho-phonemic correspondences (e.g., g in *rouge* [ʁuʒ] – "red"). A third writing task elicited four sentences that were seven to 11 words in length, for a total of 35 words in sentences. The final score was calculated as the percentage of words spelled completely and correctly out of a total of 60.

Results and Discussion

Table 2 describes the children's performance, in first grade, on the four subtests of the PHOPHLO. Speech perception performance, expressed as perception trials correct, ranged from random selection of response alternatives to perfect accuracy. Speech production accuracy, presented as percent consonants correct, was very high on average—as is expected for French—but some children scored very far below the mean. Phonological awareness, shown as number of correct items out of 14, also ranged from random guessing to perfect performance. The morphology test resulted in the full range of possible scores from 0 to 10 items produced completely correct. Therefore, with the exception of the speech production test, the effective floor and ceiling was observed in the children's responding but the mean scores were not at floor or ceiling.

Table 3 describes the performance of the 78 children who wrote the spelling test at the end of second grade. In this case, some children achieved a perfect score on one or more subtests but no child achieved a perfect total score of 60 points. All children were able to spell some non-words and real words correctly. The descriptive data presented in Tables 2 and 3 were used to define pass and fail criteria for PHOPHLO subtests and for BELO performance specifically for this sample, so as to take into account the particular characteristics of this sample including demographics, varied language background—more specifically Quebec

Table 2. Children's Performance in First Grade on PHOPHLO Subtests (*n* = 91)

Game	Construct	Min	Max	*M*	*SD*	Cut-Off Score
Écoute	Speech perception	50	100	91.10	13.94	80
Qu'est-ce que c'est?	Speech production	78	100	96.94	3.97	91
Ils aiment quoi?	Phonological awareness	5	14	11.95	2.14	10
Qu'est-ce qu'ils font?	Morphology production	0	10	7.00	2.99	3

Note. The four games comprise the screening test *Prédiction des Habiletés Orthographiques Par des Habiletés Langage Oral* (PHOPHLO), with each scored as follows: Speech Perception was scored as percent correct over 10 items; Speech Production was scored as percent correct over 10 words and 36 consonants; Phonological Awareness was scored as number correct over 14 items; Morphology Production was scored as number correct over 10 items (excluding 20 practice items overall). The cut-off scores are approximately 1.25 standard deviations below the mean, with rounding and some adjustments for skewed distributions.

Table 3. Children's Performance in Second Grade on BELO Subtests (*n* = 78)

BELO Subtest	Min	Max	*M*	*SD*	*M* - 1.25 *SD*
Nonwords	3	10	8.60	1.26	7.03
Words	2	15	10.88	3.07	7.04
Words in Sentences	9	35	28.15	4.68	22.30
Total Score	14	59	47.55	8.23	37.26

Note. The subtests comprise the spelling test *Batterie d'évaluation de lecture et d'orthographe* (BELO; George & Pech-Georgel, 2006).

French as first or second language—and the literacy teaching practices in the province. Subsequently, we examined BELO performance as a function of PHOPHLO performance more directly. In Table 4, number of children who passed or failed each PHOPHLO subtest is shown along with the corresponding mean score, using the cut-off score for each PHOPHLO subtest as shown in Table 2 (children who obtained a score below the cut-off failed the subtest). Ultimately 68 children passed the PHOPHLO screen (i.e., passed at least three subtests) and 10 children failed (i.e., failed two or more subtests).

Table 4 shows the mean and the standard deviation of the BELO score for the children who passed the PHOPHLO and the children who failed the PHOPHLO in first grade, with a total score of 38 (approximately -1.25 *SD* below the mean) being the cut-off for passing the spelling test (in other words, all children who scored 38 or above passed and all children who scored 37 or below failed). The risk of significantly poor spelling performance at the end of second grade, given poor PHOPHLO performance at the beginning of first grade, is shown. For example, the last row of Table 4 indicates that 68 children passed the PHOPHLO in first grade, achieving a mean score of 49 on the BELO in second grade with only 3% of this group failing the BELO. In other words, two children who passed the PHOPHLO in first grade failed the BELO in second grade; in contrast, 10 children failed the PHOPHLO in first grade and five of these 10 (50%) also failed BELO in second grade.

As indicated in Table 4, BELO performance is lower for children who failed than for children who passed the PHOPHLO subtests. The mean differences were submitted to nonparametric randomization tests

Table 4. Second-Grade BELO Performance as a Function of Passing or Failing PHOPHLO in First Grade

Game	Construct	Spelling Performance (BELO Scores) in Second Grade							
		Pass PHOPHLO				Fail PHOPHLO			
		M	*SD*	*n**	Fail BELO *n,%*	*M*	*SD*	*n*†	Fail BELO *n,%*
Écoute	Speech perception	48.17	6.97	70	4, 6%	42.13	15.12	8	3, 38%
Qu'est-ce que c'est?	Speech production	48.40	7.13	68	5, 7%	41.80	12.63	10	2, 20%
Ils aiment quoi?	Phonological awareness	48.35	8.15	65	4, 6%	43.54	7.74	13	3, 23%
Qu'est-ce qu'ils font?	Morphology production	48.97	6.86	63	3, 5%	41.60	10.83	15	4, 27%
PHOPHLO	Fail 2 or more subtests	49.01	3.37	68	2, 3%	37.60	12.32	10	5, 50%

Note. PHOPHLO = *Prédiction des Habiletés Orthographiques Par des Habiletés Langage Oral;* BELO = *Batterie d'évaluation de lecture et d'orthographe;* * *this column indicates the number (*n*) of students who passed the PHOPHLO (sub)test and is the denominator for the percentage (%) of students who failed the BELO, given that they passed the PHOPHLO (sub)test in first grade; †this column indicates the number (*n*) of students who failed the PHOPHLO (sub)test and is the denominator for the percentage (%) of students who failed the BELO, given that they failed the PHOPHLO (sub)test in first grade.

(Edgington & Onghena, 2007) because the sample sizes were very different and therefore the assumption of homogeneity of variance was not met, precluding parametric tests. For the subtest *Écoute*, $t = -0.641$, $p = .26$, $d = 0.374$; for *Qu'est-ce que c'est?*, $t = -2.44$, $p = .01$, $d = 0.413$; for *Ils aiment quoi?*, $t = -1.96$, $p = .04$, $d = 0.298$; for *Qu'est-ce qu'ils font?*, $t = -3.31$, $p = .002$, $d = 0.476$; and for the PHOPHLO, $t = -4.60$, $p < .001$, $d = 0.779$. Therefore, it can be seen that the mean differences in BELO performance were statistically significant for three subtests: those targeting speech production, phonological awareness, and morphology production. The largest effect size was obtained when total screening test performance was taken into account.

The information about the probability of passing the BELO, given a failure on the PHOPHLO approximately 18 months prior, yields a sensitivity of 71% (i.e., proportion of true positives identified) and specificity of 93% (i.e., proportion of true negatives identified) for the PHOPHLO as a screen for spelling difficulties in this sample. The data for these calculations are provided in Table 5, along with the likelihood ratio, indicating that a second-grade poor speller was 10 times more likely to have failed the PHOPHLO in first grade than a good speller.

Some details about the children who failed either the PHOPHLO screening in first grade or the BELO spelling test in second grade are shown in Table 6. It is instructive to consider the cases of successful and unsuccessful prediction separately, especially in relation to the language background of the students. Although this group of children is very small, some patterns in these data inform hypotheses for future research.

Considering the children who failed the PHOPHLO and the BELO (the true positives), three of the children demonstrated difficulties with phonological representations, specifically failing the speech perception test along with either the speech production or phonological awareness subtests. The remaining two children had difficulty with phonological and non-phonological language skills, that is, the phonological awareness and morphology production subtests of the PHOPHLO. Three of the five children were male and all were monolingual speakers of French. The parents of three children reported concerns that the children might be at risk for dyslexia due to a family history, and a fourth child had reported issues with conductive hearing loss. In first grade, the average teacher rating of risk for future writing problems

Table 5. Performance of the PHOPHLO as a Screening Tool

	Fail BELO	Pass BELO	Row Totals
Fail PHOPHLO	5	5	10
Pass PHOPHLO	2	66	68
Column Totals	7	71	78
Likelihood Ratio	0.71	0.07	10.14

Note. PHOPHLO = *Prédiction des Habiletés Orthographiques Par des Habiletés Langage Oral;* BELO = *Batterie d'évaluation de lecture et d'orthographe.*

Table 6. Test Scores of Children who Failed the BELO or the PHOPHLO

BELO Status	BELO Total	Speech Perception	Speech Production	Phonological Awareness	Morphology Production	PHOPHLO Status	Language
Fail	32	100	100	7	0	Fail	ML
Fail	30	90	100	7	1	Fail	ML
Fail	37	60	100	9	8	Fail	ML
Fail	14	70	86	10	0	Fail	ML
Fail	26	50	78	11	10	Fail	ML
Fail	32	100	100	10	0	Pass	ML
Fail	30	100	94	11	9	Pass	ML
Pass	52	50	100	12	2	Fail	ML
Pass	49	50	91	8	0	Fail	BL
Pass	43	90	89	7	0	Fail	BL
Pass	52	100	89	8	1	Fail	BL
Pass	41	90	94	8	0	Fail	BL

Note. PHOPHLO = *Prédiction des Habiletés Orthographiques Par des Habiletés Langage Oral;* BELO = *Batterie d'évaluation de lecture et d'orthographe;* ML = monolingual; BL = bilingual; blue shading indicates that the child failed the subtest by scoring below the cut-off scores shown in Table 2.

was 2.8, higher than the mean rating of 1.6 for monolingual children in this study (see Kolne et al., 2016 for details of the teacher ratings).

Two other children failed the BELO in second grade despite passing the PHOPHLO in first grade. The parents of these male children reported concerns about their child, specifically a history of language comprehension problems that were treated in preschool in one case and significant concerns about social problems with peers and aggressive behavior with respect to the other child. The teacher ratings of concern about future writing difficulties were relatively high at 3.5.

Five other children failed the PHOPHLO but passed the BELO in second grade: all five failed the morphology production subtest of the PHOPHLO and four failed the phonological awareness subtest. Four of the five children were drawn from the BL subsample, in other words speaking a language that was not French at home. The parents of these five female children reported no concerns about their development. Their teachers provided a mean rating of concern about their future writing skills of 3.2, however—higher than the mean rating for BL children of 2.0.

Table 6 shows that the failures of sensitivity occurred within the subgroup of ML children such that two of seven MLs who failed the BELO were not identified by the PHOPHLO. These two children who proved to be poor spellers after passing the PHOPHLO screen highlight the fact that this screening test does not measure the child's performance in all domains of knowledge that are known to predict literacy outcomes.

The failures of specificity were largely due to BL language exposure. It seems that children who are not fully competent in their French language skills at school entry can achieve good spelling skills by second grade. This may occur because ML and BL children have received similar exposures to written language instruction whereas these two groups have had quite different experiences in the oral language domain. It is not certain that the BL children who failed the PHOPHLO at school entry will be successful on the provincial literacy exam in fourth grade given that it requires integration of written language skills across a variety of areas including reading comprehension as well as overall coherence, syntax, punctuation, spelling, and grammar when writing a narrative.

General Discussion

Oral language skills are readily observable at or before the onset of formal reading and writing instruction and may predict the child's future response to formal instruction in school, independently of variations in access to direct literacy instruction in the home or preschool environment. Therefore, we conducted a two-phase study to develop a screening procedure that is focused on oral language abilities for the purpose of identifying children who may struggle to learn to read and spell in the early school years. In this study, we screened children's speech perception, speech production, phonological awareness, and morphology production abilities at entry to first grade and predicted spelling skills at the end of second grade. The results of the study will be discussed in relation to our objectives first. Subsequently, the limitations of the study will be discussed in detail.

Development and performance of the PHOPHLO

Objective 1. Given that oral language abilities predict the acquisition of literacy skills in general and spelling in particular (Pennala et al., 2010; Speece et al., 1999), the first objective was to examine the role of speech perception, speech production, phonological awareness, and morphology production in differentiating children who should differ in literacy skills. In Phase I, we tested 61 children with the full versions of our tests, and found that certain subtests differentiated children with higher versus lower performance overall. Phonological awareness performance was especially discriminating but speech perception and production were also effective. ML versus BL children performed differently on the morphology production test. Therefore the Phase I results suggested that it was prudent to continue the development of the screener with all four constructs represented.

Objective 2. A second important objective in the first phase was to reduce the total number of items to create a screening test that could be administered in a much shorter period of time while covering the same four constructs. Ultimately the number of items was reduced from 152 to 64 items in a screener that contained 44 test items and 20 practice items. The correlation between the shorter screening test and the longer test battery was .89. However, future studies are necessary to establish the reliability of this screening test within and across screeners, especially those with different training and preparation, and in varied school environments.

Objective 3. The primary objective of Phase II was to determine the sensitivity and the specificity of the screener to predict spelling performance at the end of second grade. For our sample of suburban Quebec children in which a large proportion were bilingual, the PHOPHLO proved to

have a sensitivity of 71% and a specificity of 93% as a screen for poor spelling abilities. Clearly, further study is required to replicate this result as will be discussed further.

Objective 4. Questions about the contribution of the four constructs to the utility of the screening test continued into the second phase of the study. PHOPHLO performance in first grade was associated with BELO performance in second grade, especially with respect to speech production, phonological awareness, and morphology production. The subtest targeting speech perception did not differentiate children with respect to mean BELO score. It is possible that this subtest is particularly vulnerable to poor performance due to extraneous variables that do not elevate the child's risk of spelling difficulties; these may include a noisy environment, poor comprehension of instructions, poor attention, or transient hearing problems on the part of the child. On the other hand, examination of the individual child data in Table 6 suggests that, in some cases, poor speech perception performance may combine with poor speech production and phonological awareness skills to indicate a generalized problem with phonological representations or phonological processing. Ramus, Marshall, Rosen, & Van der Lely (2013) have suggested that children with dyslexia fall into two profiles: those who have difficulty with phonological representations (as revealed by speech perception and production tasks), and those who have difficulty with phonological and non-phonological language skills (as revealed by phonological awareness and language production tasks). Therefore, it seems worthwhile to continue research with all four subtests so as to accumulate data from a larger group of true positives.

Limitations and future directions

A significant limitation of this study is the small sample size for assessing the predictive validity of the PHOPHLO. Clearly replication samples are required to confirm our estimate of the sensitivity and specificity of the PHOPHLO as a screen for spelling impairments in second grade. We feel that the mixed language background of our sample is a strength of the study given the increasingly multilingual and multicultural characteristic of the school population. However, a larger sample of children for validation of the screening tool would provide greater confidence in the sensitivity and specificity results, while permitting an exploration of differences in predictive accuracy within different subsets of the validation sample. Certainly, exploration of differences across ML versus BL groups would require a very large sample. A first priority would be to cross-validate the results with a larger sample of children with similar composition to that described here. Subsequently, follow-up studies with more varied samples, including children with lower maternal education, for example, would be advisable.

Another subgroup analysis that would be enabled by a larger validation sample would concern the emergence of possible gender differences in literacy skills during the primary grades. Although significant gender differences are observed on the obligatory written language competency exam in Quebec, we did not observe any gender differences in PHOPHLO performance at school entry. Our sample was too small and unbalanced to explore this issue further. Limbrick, Wheldall, and Madelaine (2012) found that boys and girls do not differ in any aspect of literacy performance in the early school grades and suggested that gender differences emerge over time because of an increasing gap between school expectations and boys' behaviour.

A second limitation of the study, also related to its scope, is the restriction of the predictor and outcome variables to a narrow range, specifically oral language predictors and spelling as the outcome variable. With respect to the predictor variables it is known that there are other types of predictors that are useful as predictors of literacy outcomes. For younger children, print concepts in general and letter knowledge especially is an effective predictor (Erdos, Genesee, Savage, & Haigh, 2011; Storch & Whitehurst, 2002). For older children, orthographic knowledge is another important correlate of reading and spelling abilities (Binamé & Poncelet, 2016; Bourgoin, 2014; Commissaire, Pasquarella, Chen, & Deacon, 2014; Cunningham, Perry, & Stanovich, 2001). Stanké, Flessas, and Ska (2008) describe tests of orthographic processing that are available for testing French-speaking children. A necessary future step would be to determine if the PHOPHLO provides any predictive value over and above that offered by screening tests such as the *Outil de dépistage d'élèves à risque de présenter des difficultés d'apprentisage du langage écrit* (ODLÉ; Stanké & Flessas, 2013). The ODLÉ assesses phonological awareness, visual memory, and orthographic memory, and has been normed on large samples of French speaking children from kindergarten and first-grade classes in Quebec. A combination of oral language and orthographic screening might offer improved sensitivity over oral language screening alone. This raises another limitation of our study, and that is the single time point for screening being first-grade entry. However, a more adequate screening protocol would likely involve layered screenings, for example oral language screening in kindergarten followed by orthographic screening in first grade (after the children have received systematic exposure to written language instruction).

An investigation of the effectiveness of PHOPHLO as a screening tool when used at an earlier age and in the context of a more comprehensive screening protocol would be desirable.

Regarding the outcome measure, spelling was selected as an early indicator of writing abilities that are causing concern on the province-wide competency exam. However, oral language abilities are known to predict many aspects of literacy and therefore a more extensive outcome battery including real word and non-word reading accuracy in addition to spelling would be desirable. Linking early screening to actual performance on the obligatory writing competency examination would also be particularly valuable.

A final limitation of the study is the lack of concurrent validation of the PHOPHLO screening test with another measure of oral language abilities such as the *Petite Évac (Épreuve Verbale d'Aptitudes Cognitives pour les petits de 3 à 9 ans;* Lussier, Flessas, & Stanké, 2003). An assessment of the performance of the PHOPHLO in relation to the *Petite Évac* with respect to concurrent and predictive validity would be informative, not only as an indicator of the validity of the PHOPHLO but as an examination of relative efficiency. The administration time for the PHOPHLO is one half to one quarter the time required for the *Petite Évac*; furthermore, the digital implementation of the PHOPHLO permits administration and interpretation by paraprofessionals. The PHOPHLO could be a useful tool for identifying children who require more extensive testing by speech-language pathologists. We note that several children in our sample were suspected to have language impairments but were awaiting speech-language assessments throughout the course of the study.

Conclusion

We have developed a digital tool that targets children's oral language abilities in four areas of language function, specifically speech perception, speech production, phonological awareness, and morphology production. Our study is unique in the inclusion of an authentic validation sample, including monolingual speakers of French, simultaneous bilingual speakers of French, and children who were first exposed to French upon preschool entry. In a preliminary investigation we have shown that, when administered early in first grade, the screen identified children who were likely to fail a spelling test at the end of second grade with a sensitivity of 71% and specificity of 93%. We have discussed important future directions for this research, given the limitations of our small sample and the need to further investigate the role of oral language skills at school entry in the emergence of written language competence in French-speaking elementary school children in Canada.

References

Alamargot, D. (2007). Développement de la mémoire: impact sur l'apprentissage de la production écrite. *Entretiens d'orthophonie Bichat*, 210–228.

Anthony, J. L., Lonigan, C. J., Driscoll, K., Philips, B. M., & Burgess, S. R. (2003). Phonological sensitivity: A quasi-parallel progression of word structure units and cognitive operations. *Reading Research Quarterly, 38*(4), 470–487. doi: 10.1598/RRQ.38.4.3

Bédard, D., Marquis, A., Royle, P., Gonnerman, L. M., & Rvachew, S. (2013, October). *Comparing oral and written morphosyntax in monolingual and multilingual children: A longitudinal study.* Paper presented at the International Conference on Multilingualism: Linguistic Challenges and Neurocognitive Mechanisms, McGill University, Montréal, Canada.

Béchara, J. (2015). *Évaluation de la production morphosyntaxique chez un enfant québécois francophone porteur d'implant cochléaire* (Master's thesis). Université de Montréal, Montréal, Canada.

Berninger, V. W., Nielsen, K. H., Abbott, R. D., Wijsman, E., & Raskind, W. (2008). Writing problems in developmental dyslexia: Under-recognized and under-treated. *Journal of School Psychology, 46*, 1–21. doi: 10.1016/j.jsp.2006.11.008

Binamé, F., & Poncelet, M. (2016). The development of the abilities to acquire novel detailed orthographic representations and maintain them in long-term memory. *Journal of Experimental Child Psychology, 143*, 14–33. doi: 10.1016/j.jecp.2015.10.010

Bird, J., Bishop, D. V. M., & Freeman, N. H. (1995). Phonological awareness and literacy development in children with expressive phonological impairments. *Journal of Speech and Hearing Research, 38*, 446–462. doi:10.1044/jshr.3802.446

Boets, B., Wouters, J., van Wieringen, A., De Smedt, B., & Ghesquière, P. (2008). Modelling relations between sensory processing, speech perception, orthographic and phonological ability, and literacy achievement. *Brain and Language, 106*(1), 29–40. doi: 10.1044/jshr.3802.446

Bourassa, D. C., & Treiman, R. (2001). Spelling development and disability: The importance of linguistic factors. *Language, Speech, and Hearing Services in Schools, 32*, 172–181. doi: 10.1044/0161-1461(2001/016)

Bourgoin, R. (2014). The predictive effects of L1 and L2 early literacy indicators on reading in French immersion. *Canadian Modern Language Review, 70*, 1–26. doi: 10.3138/cmlr.2346

Brosseau-Lapré, F. & Rvachew, S. (2008). Test de Conscience Phonologique Préscolaire [Computer Software]. Montréal, QC: Author.

Brosseau-Lapré, F., & Rvachew, S. (2014). Cross-linguistic comparison of speech errors produced by English- and French-speaking preschool age children with developmental phonological disorders. *International Journal of Speech-Language Pathology, 16*(2), 98–108. doi: 10.3109/17549507.2013.794863

Brosseau-Lapré, F., & Rvachew, S. (2017). Underlying manifestations of developmental phonological disorders in French-speaking pre-schoolers. *Journal of Child Language, 44,* 1337-1361. doi: 10.1017/S0305000916000556

Burton, R. F. (2001). Do item-discrimination indices really help us to improve our tests? *Assessment & Evaluation in Higher Education, 26*, 213–220. doi: 10.1080/02602930120052378

Caravolas, M., & Bruck, M. (2000). Vowel categorization skill and its relationship to early literacy skills among first-grade Québec-French children. *Journal of Experimental Child Psychology, 76*(3), 190–221. doi: 10.1006/jecp.1999.2547

Charest, D. (2010). *Évaluation de programme : plan d'action pour l'amélioration du français (premier rapport d'étape)*. Retrieved from Gouvernement du Québec, Ministère de l'Éducation, Loisir et du Sport website: http://www.

education.gouv.qc.ca/references/publications/resultats-de-la-recherche/detail/article/evaluation-de-programme-plan-daction-pour-lamelioration-du-francais-premier-rapport-detape/

Commissaire, E., Pasquarella, A., Chen, X., & Deacon, S. H. (2014). The development of orthographic processing skills in children in early French immersion programs. *Written Language & Literacy, 17,* 16–39. doi: 10.1075/wll.17.1.02com

Connelly, V., Campbell, S., MacLean, M., & Barnes, J. (2006). Contribution of lower order skills to the written composition of college students with and without dyslexia. *Developmental Neuropsychology, 29*, 175–196. doi: 10.1207/s15326942dn2901_9

Cooper, D. H., Roth, F. P., Speece, D. L., & Schatschneider, C. (2002). The contribution of oral language skills to the development of phonological awareness. *Applied Psycholinguistics, 23*, 399–416. doi: 10.1017.S0142716402003053

Cunningham, A. E., Perry, K. E., & Stanovich, K. E. (2001). Converging evidence for the concept of orthographic processing. *Reading and Writing: An Interdisciplinary Journal, 14*, 549–568. doi 10.1023/A:1011100226798

Dewalt, D. A., Berkman, N. D., Sheridan, S., Lohr, K. N., & Pignone, M. P. (2004). Literacy and health outcomes: A systematic review of the literature. *Journal of General Internal Medicine, 19*, 1228–1239. doi: 10.1111/j.1525-1497.2004.40153.x

Dickinson, D. K., McCabe, A., Anastasopoulos, L., Peisner-Feinberg, E. S., & Poe, M. D. (2003). The comprehensive language approach to early literacy: The interrelationships among vocabulary, phonological sensitivity, and print knowledge among preschool-aged children. *Journal of Educational Psychology, 95*(3), 465–481. doi: 10.1037/0022-0663.95.3.465

Dickinson, D. K., & Porsche, M. V. (2011). Relation between language experiences in preschool classrooms and children's kindergarten and fourth-grade language and reading abilities. *Child Development, 82*(3), 870–886. doi: 10.1111/j.1467-8624.2011.01576.x

Duncan, L., Colé, P., & Casalis, S. (2009). Early metalinguistic awareness of derivational morphology: Observations from a comparison of English and French. *Applied Psycholinguistics, 30*, 405–440. doi: 10.1017/S0142716409090213

Durand, V. N., Loe, I. M., Yeatman, J. D., & Feldman, H. M. (2013). Effects of early language, speech, and cognition on later reading: A mediation analysis. *Frontiers in Psychology, 4*, 1-11. doi: 10.3389/fpsyg.2013.00586

Edgington, E. S., & Onghena, P. (2007). *Randomization tests*. New York, NY: Chapman & Hall.

Ehri, L. C., Wilce, L. S., & Taylor, B. B. (1987). Children's categorization of short vowels and the influence of spellings. *Merrill-Palmer Quarterly, 33*, 393–421. Retrieved from http://www.jstor.org/stable/23086540

Erdos, C., Genesee, F., Savage, R., & Haigh, C. A. (2011). Individual differences in second language reading outcomes. *International Journal of Bilingualism, 15*(1), 3–25. doi: 10.1177/1367006910371022

Fortin, L., Royer, É., Potvin, P., Marcotte, D., & Yergeau, É. (2004). La prédiction du risque de décrochage scolaire au secondaire : facteurs personnels, familiaux et scolaires. *Revue canadienne des sciences du comportement, 36*, 219–231. doi: 10.1037/h0087232

Gelman, A., & Park, D. K. (2009). Splitting a predictor at the upper quarter or third and the lower quarter or third. *The American Statistician, 63*, 1–8. doi: /10.1198/tast.2009.0001

George, F., & Pech-Georgel, C. (2006). *BELO – Batterie d'évaluation de lecture et d'orthographe*. Dardilly, France: Éditions Solal.

Gouvernement du Québec. (2006). *Bilan de l'application du programme de formation de l'école québécoise - enseignement primaire : Recommandations de la Table de pilotage du renouveau pédagogique*. Retrieved from http://www.mels.gouv.qc.ca/renouveau/pdf/Recommandations.pdf

Graham, S., & Santangelo, T. (2014). Does spelling instruction make students better spellers, readers, and writers? A meta-analytic review. *Reading and Writing, 27*(9), 1703–1743. doi: 10.1007/s11145-014-9517-0

Griffin, T. M., Hemphill, L., Camp, L., & Palmer Wolf, D. (2004). Oral discourse in the preschool years and later literacy skills. *First Language, 24*, 123–146. doi: 10.1177/0142723704042369

Hernandez, D. J. (2011). *Double jeopardy: How third grade reading skills and poverty influence high school graduation*. Baltimore, MD: Annie E. Casey Foundation.

Holm, A., Farrier, F., & Dodd, B. (2008). Phonological awareness, reading accuracy and spelling ability of children with inconsistent phonological disorder. *International Journal of Language & Communication Disorders, 43*, 300–322. doi: 10.1080/13682820701445032

Hulme, C., Nash, H. M., Gooch, D., Lervag, A., & Snowling, M. J. (2015). The foundations of literacy development in children at familial risk of dyslexia. *Psychological Science, 26*, 1877–1886. doi: 10.1080/13682820701445032

Jalbert, P. (2007). *L'épreuve obligatoire d'écriture de la fin du troisième cycle du primaire en français, langue d'enseignement : Comparaison des résultats de 2000 et 2005*. Retrieved from Gouvernement du Québec, Ministère de l'Éducation, Loisir et du Sport website: http://www.education.gouv.qc.ca/fileadmin/site_web/documents/dpse/evaluation/Ecriture3eCyclePrimFLE.pdf

Justice, L. M. (2006). Evidence-based practice, response to intervention, and the prevention of reading difficulties. *Language, Speech & Hearing Services in Schools, 37*, 284–297. doi:10.1044/0161-1461(2006/033)

Kelley, T. L. (1939). The selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology, 30*, 17–24.

Kolne, K., Gonnerman, L., Marquis, A., Royle, P., & Rvachew, S. (2016). Teacher predictions of children's spelling ability: What are they based on and how good are they? *Language and Literacy, 18*(1), 7198. Retrieved from https://journals.library.ualberta.ca/langandlit/index.php/langandlit/article/viewFile/22994/20260

Kresh, S. (2008). What's regular (and what's not) in French-speaking children's processing of inflectional morphology? In A. Gavarró and M. J. Freitas (Eds.), *Language Acquisition and Development. Proceedings of GALA 2007*. Newcastle, United Kingdom: Cambridge Scholars Publishing.

Labrecque, M., Chuy, M., Brochu, P., & Houme, K. (2012). PIRLS 2011. *Le contexte au Canada: Résultats canadiens du Programme international de recherche en lecture scolaire.* Retrieved from Conseil des ministres de l'Éducation (Canada) website: http://cmec.ca/Publications/Lists/Publications/Attachments/294/PIRLS_2011_FR.pdf

Lefavrais, P. (2005). *L'Alouette-R*. Paris, France: Les Editions du Centre de Psychologie Appliquée.

Lefebvre, P., Trudeau, N., & Sutton, A. (2008). Prevention of reading and writing difficulties: A preliminary study of the practices of Canadian speech-language pathologists. *Canadian Journal of Speech-Language Pathology and Audiology, 32*, 74–84.

Legendre, G., Goyet, L., Barrière, I., Kresh, S., & Nazzi, T. (2009, November). *Sensitivity to irregular French subject-verb agreement at 18 months: Evidence from the Head-turn Preference Procedure*. Paper presented at the 34th annual Boston University Conference on Language Development, Boston, MA.

Lewis, B. A., Freebairn, L. A., & Taylor, H. G. (2002). Correlates of spelling abilities in children with early speech sound disorders. *Reading and Writing: An Interdisciplinary Journal, 15*, 389–407. doi: 10.1023/A:1015237202592

Limbrick, L., Wheldall, K., & Madelaine, A. (2012). Reading and related skills in the early school years: Are boys really more likely to struggle? *International Journal of Disability, Development and Education, 59*(4), 341–358. doi: 10.1080/1034912X.2012.723939

Lussier, F., Flessas, J., & Stanké, B. (2003). *Épreuve Verbale d'Aptitudes Cognitives pour les petits de 3 à 9 ans (La petite Évac)*. Montréal, Canada : Centre d'Évaluation Neuropsychologique et d'Orientation Pédagogique FL.

Marcus, E. N. (2006). The silent epidemic - The health effects of literacy. *New England Journal of Medicine, 355*(4), 339–342. doi: 10.1056/NEJMp058328

Marquis, A., Royle, A., Gonnerman, L. M., & Rvachew, S. (2012). La conjugaison du verbe en début de scolarisation. *Travaux interdisciplinaires sur la parole et le langage, 28*. Retrieved from http://tipa.revues.org/201

Nation, K., & Snowling, M. J. (2004). Beyond phonological skills: Broader language skills contribute to the development of reading. *Journal of Research in Reading, 27*(4), 342–356. doi: 10.1111/j.1467-9817.2004.00238.x

Organization for Economic Cooperation and Development [OECD]. (2011). *Literacy for life: Further results from the Adult Literacy and Life Skills Survey.* Ottawa, Canada: OECD Publishing.

Overby, M., & Bernthal, J. E. (2008, November). *A study of speech sound perception and phonological spelling in children with SSD.* Paper presented at the American Speech-Language-Hearing Association, Chicago, IL.

Overby, M. S., Masterson, J. J., & Preston, J. L. (2015). Preliteracy speech sound production skill and linguistic characteristics of grade 3 spellings: A study using the Templin archive. *Journal of Speech, Language, and Hearing Research,* 1–16. doi: 10.1044/2015_JSLHR-S-14-0276

Pacton, S., & Deacon, S. H. (2008). The timing and mechanisms of children's use of morphological information in spelling: A review of evidence from English and French. *Cognitive Development, 23*, 339–359. doi: 10.1016/j.cogdev.2007.09.004

Paul, M. (2009). *Predictors of consonant development and the development of a test of French phonology* (Master's thesis). McGill University, Montréal, Canada.

Pech-Georgel, C., & George, F. (2010). Batterie d'évaluation des troubles du langage écrit adaptée aux lycéens et adultes dyslexiques. *Développements, 3*(6), 27–34.

Pennala, R., Eklund, K., Hämäläinen, J., Richardson, U., Martin, M., Leiwo, M., ... Lyytinen, H. (2010). Perception of phonemic length and its relation to reading and spelling skills in children with family risk for dyslexia in the first three grades of school. *Journal of Speech, Language & Hearing Research, 53*(3), 710–724. doi: 10.1044/1092-4388(2009/08-0133)

Puranik, C. S., Petcher, Y., Al Otaiba, S., Catts, H. W., & Lonigan, C. J. (2008). Development of oral reading fluency in children with speech or language impairments: A growth curve analysis. *Journal of Learning Disabilities, 41*, 545–560. doi: 10.1177/0022219408317858

Ramus, F., Marshall, C., Rosen, S., & Van der Lely, H. K. J. (2013). Phonological deficits in specific language impairment and developmental dyslexia: A multidimensional model. *Brain, 136*, 630–645. doi: 10.1093/brain/aws356

Roth, F. P., & Baden, B. (2001). Investing in emergent literacy intervention: A key role for speech-language pathologists. *Seminars in Speech and Language, 22*(3), 163–174. doi: 10.1055/s-2001-16143

Royle, P. (2007). Variable effects of morphology and frequency on inflection patterns of French preschoolers. *The Mental Lexicon Journal, 2*, 103–125.

Royle, P., & Thordardottir, E. (2008). Elicitation of the passé-composé in French preschoolers with and without specific language impairment. *Applied Psycholinguistics, 29*(3), 341–365. doi: 10.1017/S0142716408080168

Rvachew, S. (2007). Phonological processing and reading in children with speech sound disorders. *American Journal of Speech-Language Pathology, 16*, 260–270. doi:10.1044/1058-0360(2007/030)

Rvachew, S. (2009). Speech Assessment and Interactive Learning System (Version 2) [Software]. Montréal, Canada: Author.

Rvachew, S., & Brosseau-Lapré, F. (2015). A randomized trial of twelve week interventions for the treatment of developmental phonological disorder in francophone children. *American Journal of Speech-Language Pathology, 24*, 637-658. doi: 10.1044/2015_AJSLP-14-0056

Rvachew, S., & Grawburg, M. (2006). Correlates of phonological awareness in preschoolers with speech sound disorders. *Journal of Speech, Language, and Hearing Research, 49*, 74–87. doi:10.1044/1092-4388(2006/006)

Rvachew, S., Marquis, A., Brosseau-Lapré, F., Paul, M., Royle, P., & Gonnerman, L. M. (2013). Speech articulation performance of francophone children in the early school years: Norming of the Test de Dépistage Francophone de Phonologie. *Clinical Linguistics & Phonetics, 27*(12), 950–968. doi: 10.3109/02699206.2013.830149

Schneider, W., Roth, E., & Ennemoser, M. (2000). Training phonological skills and letter knowledge in children at risk for dyslexia: A comparison of three kindergarten intervention programs. *Journal of Educational Psychology, 92*, 284–295.

Sénéchal, M., Basque, M. T., & Leclaire, T. (2006). Morphological knowledge as revealed in children's spelling accuracy and reports of spelling strategies. *Journal of Experimental Child Psychology, 95*, 231–254. doi: 10.1016/j.jecp.2006.05.003

Sénéchal, M., & LeFevre, J. (2002). Parental involvement in the development of children's reading: A five-year longitudinal study. *Child Development, 73*(2), 445–460. doi: 10.1111/1467-8624.00417/pdf

Snowling, M. J., & Melby-Lervåg, M. (2016). Oral language deficits in familial dyslexia: A meta-analysis and review. *Psychological Bulletin, 142*(5), 498–545. doi: 10.1037/bul0000037

Speece, D. L., Roth, F. P., Cooper, D. H., & de la Paz, S. (1999). The relevance of oral language skills to early literacy: A multivariate analysis. *Applied Psycholinguistics, 20*, 167–190.

Stanké, B., & Flessas, J. (2013). *Outil de dépistage d'élèves à risque de présenter des difficultés d'apprentisage du langage écrit (ODLÉ)*. Montréal, Canada : IRP.

Stanké, B., Flessas, J., & Ska, B. (2008). Le rôle de la mémoire lexicale orthographique dans l'acquisition des connaissances orthographiques des enfants de maternelle 5 ans. *A.N.A.E. (Approche Neuropsychologique des Apprentissages chez l'Enfant), 100*, 326–335.

Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology, 38*(6), 934–947. doi: 10.1037/0012-1649.38.6.934

Stothard, S. E., Snowling, M. J., Bishop, D. V., Chipchase, B. B., & Kaplan, C. A. (1998). Language-impaired preschoolers: A follow-up into adolescence. *Journal of Speech, Language, and Hearing Research, 41*(2), 407–418. doi:10.1044/jslhr.4102.407

Thordardottir, E. T., & Namazi, M. (2007). Specific language impairment in French-speaking children: beyond grammatical morphology. *Journal of Speech, Language, and Hearing Research, 50*(3), 698–715. doi: 10.1044/1092-4388(2007/049)

Weiser, B., & Mathes, P. (2011). Using encoding instruction to improve the reading and spelling performances of elementary students at risk for literacy difficulties: A best-evidence synthesis. *Review of Educational Research, 81*, 170–200. doi: 10.3102/0034654310396719

Wolter, J. A., Wood, A., & D'zatko, K. W. (2009). The influence of morphological awareness on the literacy development of first-grade children. *Language, Speech and Hearing Services in Schools, 40*, 286–298. doi: 10.1044/0161-1461(2009/08-0001)

Acknowledgements

This study was conducted with support from FQRSC (2011-ER-144359, Rvachew et al.) and MITACS (MITACS IT06307, Rvachew et al.) and Subvention Institutionnelle du CRSH – UdeM (Royle & Marquis, 2014).

We thank the schools and all the children who participated in this project. We are also grateful to the research assistants who collected and coded data,

as follows: Geneviève Beauregard-Paultre, Madeleine Bellemare, Catherine Clémence, Tara Commandeur, Kendall Kolne, and Catherine Lamirande. We also acknowledge essential support from CRIM (Centre de Recherche Informatique de Montréal), which supervised the development of the PHOPHLO software.

Authors' Note

English correspondence concerning this article should be addressed to Susan Rvachew, School of Communication Sciences and Disorders, McGill University, 2001 McGill College Ave., 8th Floor, Montréal, Quebec, Canada, H3A 1G1. Email: **susan.rvachew@mcgill.ca.**

Les demandes en français au sujet de cet article doivent être adressées à Phaedra Royle, École d'orthophonie et audiologie, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, Québec, Canada. Courriel : **phaedra.royle@umontreal.ca**.

SAC
OAC

Speech-Language &
Audiology Canada

Orthophonie et
Audiologie Canada

Communicating care
La communication à coeur

613.567.9968
1.800.259.8519
1000-1 rue Nicholas St.
Ottawa ON K1N 7B7
www.sac-oac.ca | @SAC_OAC