



Revue critique de la littérature sur les qualités métrologiques du Hearing in Noise Test



A critical literature review on the psychometric properties of the Hearing in Noise Test

| MOTS CLÉS |
|---------------------------|
| HEARING IN NOISE TEST |
| <i>HINT</i> |
| PROPRIÉTÉS MÉTROLOGIQUES |
| APPLICABILITÉ |
| VALIDITÉ |
| FIDÉLITÉ |
| SENSIBILITÉ AU CHANGEMENT |
| VERSIONS LINGUISTIQUES |

Mathieu Hotton
François Bergeron

Abrégé

Le *Hearing in Noise Test (HINT)* est utilisé internationalement pour évaluer le seuil de reconnaissance de phrases (SRPh) dans le silence et le bruit. Toutefois, les qualités métrologiques des différentes versions linguistiques n’ont jamais été revues de manière exhaustive. Une revue critique de la littérature a été réalisée pour vérifier l’applicabilité, la validité, la fidélité et la sensibilité des versions linguistiques du *HINT* pour mesurer le SRPh, dans le silence et le bruit, chez des personnes ayant une acuité auditive normale ou une surdité, et pour identifier les normes disponibles. Les bases de données *Cochrane Library*, *Medline*, *Embase*, *CINAHL* et *PsycINFO* ont été consultées. Seuls les articles rédigés en anglais ou en français et publiés dans des périodiques revus par des pairs ont été inclus. La collecte de données et l’évaluation de la qualité scientifique des articles ont été réalisées à l’aide des outils de Law & MacDermid (2008). Vingt-quatre articles ont été évalués. Leur qualité scientifique était généralement limitée. La plupart des études ont démontré les validités apparente et de contenu du *HINT*; deux études se sont intéressées à sa validité de construit. Les études sur la fidélité ont rapporté des erreurs de mesure et des intervalles de confiance à 95%; la fidélité inter-listes a aussi été discutée. Trois études ont évalué la fidélité auprès de personnes ayant une surdité. Aucune donnée sur la sensibilité du *HINT* n’a été retracée. Les normes sous écouteurs apparaissent variables d’une langue à l’autre; cette variabilité est plus élevée pour le SRPh dans le silence. Les faits scientifiques disponibles suggèrent que l’applicabilité, la validité, la fidélité et la sensibilité des différentes versions linguistiques du *HINT* ne sont pas toutes bien établies. Elles demeurent particulièrement méconnues chez les personnes ayant une surdité, qui constituent pourtant la population cible de ce test. Il s’agit d’une lacune importante qui devrait faire l’objet de recherches futures.

Abstract

The *Hearing in Noise Test (HINT)* is used internationally to measure the sentence recognition threshold (SRT) in quiet and in noise. However, the psychometric properties of the various linguistic versions have never been exhaustively reviewed. A critical review of the literature was completed to check the applicability, the validity, the reliability and the sensitivity of the linguistic versions of the *HINT* to measure the SRT in quiet and in noise in people with normal or impaired hearing, and to identify available norms.

Cochrane Library, *Medline*, *Embase*, *CINAHL*, and *PsycINFO* databases were consulted. Only articles in English or in French and published in peer reviewed journals were included. Data collection and the evaluation of the scientific quality of the articles were done using the tools from Law & MacDermid (2008). Twenty-four articles were evaluated. Their scientific quality was generally limited. Most of the studies demonstrated the face and contents validities of the *HINT*; two studies focused on its construct validity. Studies about reliability reported errors of measurement and confidence intervals at 95 %; interlist reliability was also discussed.

Three studies evaluated reliability with hearing-impaired people. No data on the sensitivity of *HINT* was traced. The headphone norms seem variable from one language to another ; this variability is higher for the SRT in a quiet environment. The scientific facts available suggest that the applicability, validity, reliability, and sensitivity of the various linguistic versions of the *HINT* are not all well established. They remain specifically unknown for people with a hearing loss, who are, however, the target population of this test. This is a severe gap that should be the target of future research.

Mathieu Hotton, M.O.A.
Centre Interdisciplinaire de recherche en réadaptation et intégration sociale
Département de réadaptation,
Faculté de médecine,
Université Laval
Québec (Québec)
CANADA

François Bergeron, Ph.D.
Centre Interdisciplinaire de recherche en réadaptation et intégration sociale
Département de réadaptation,
Faculté de médecine,
Université Laval
Québec (Québec)
CANADA

Diverses techniques d'audiométrie tonale et vocale sont utilisées pour procéder à l'évaluation de l'audition. L'audiométrie tonale consiste à évaluer la fonction auditive à l'aide de sons purs (tonalités de fréquence spécifique), alors que l'audiométrie vocale utilise plutôt des signaux vocaux (mots ou phrases) (Bess & Humes, 2008).

Chez une personne atteinte de surdité, l'audiométrie tonale et vocale sont aussi utilisées pour mesurer le rendement d'aides auditives (prothèse auditive ou implant cochléaire). Typiquement, des mesures audiométriques sont réalisées avant et après l'appareillage de la personne. Une amélioration de la fonction auditive à la suite de l'appareillage est interprétée comme un indice de succès de l'intervention. Lorsque l'audiométrie est utilisée à cette fin, il est recommandé d'utiliser l'audiométrie vocale, en particulier les tests vocaux réalisés en présence de bruit ambiant (McArdle & Hnath-Chilom, 2009). En effet, les résultats obtenus à ces tests seraient des indicateurs plus représentatifs du fonctionnement communicationnel quotidien puisqu'ils mettent à contribution des stimuli se rapprochant davantage de la communication humaine.

Le *Hearing in Noise Test*

Le *Hearing in Noise Test (HINT)* (Nilsson, Soli, & Sullivan, 1994) a été conçu pour mesurer le seuil de reconnaissance de phrases (SRPh) dans le silence et dans le bruit. Ce seuil correspond à la plus petite intensité sonore pour laquelle le sujet peut répéter correctement 50% des phrases présentées. La version originale anglo-américaine du *HINT* (Nilsson et al., 1994) comporte 24 listes équivalentes de 10 phrases, pouvant être combinées en 12 listes de 20 phrases, ajustées en intelligibilité et en contenu phonémique.

Le *HINT* est administré en cabine audiométrique à l'aide d'un ordinateur. La présentation des stimuli se fait via des écouteurs ou des haut-parleurs. Le SRPh est mesuré dans quatre conditions d'écoute: dans le silence, lorsque le bruit est émis en face du sujet (« bruit de face », NF), lorsque le bruit est émis à sa droite (« bruit à droite », NR) et lorsque le bruit est émis à sa gauche (« bruit à gauche », NL). Alors que les phrases sont toujours présentées en face de la personne, la provenance du bruit est variée au cours du test, qu'il soit administré à l'aide d'écouteurs ou de haut-parleurs. Le sujet doit répéter chaque phrase présentée dans son intégralité, en considérant les équivalences acceptées pour certains mots (par exemple, « Je voyage par autobus » versus « Je voyage en autobus »). Le niveau de présentation des phrases varie en fonction de la réponse du sujet: si l'item est réussi, l'intensité de la phrase suivante est

diminuée; s'il est échoué, l'intensité de la phrase suivante est augmentée. Des variations de ± 4 dB sont appliquées aux quatre premières phrases présentées au sujet alors qu'elles sont de ± 2 dB pour les phrases subséquentes. Pour la mesure du SRPh dans le bruit, le niveau de présentation du bruit est maintenu constant à 65 dB(A). Le SRPh dans le silence est exprimé en dB(A). Le SRPh dans le bruit est exprimé sous la forme d'un ratio signal/bruit en dB (S/B), correspondant à la différence entre les niveaux de présentation des phrases et du bruit (par exemple, si le niveau de présentation des phrases au SRPh est de 55 dB(A) dans 65 dB(A) de bruit, le ratio signal/bruit sera de -10 dB). Cinq scores de SRPh sont ainsi obtenus : dans le silence, dans le « bruit de face », dans le « bruit à droite », dans le « bruit à gauche » et un dernier SRPh appelé « bruit composite » (NC). Le SRPh « bruit composite » est un SRPh moyen pour les trois conditions de bruit, calculé selon la formule suivante: $NC = [(2 \times NF) + NR + NL] / 4$.

Le *HINT* est couramment utilisé pour évaluer l'audition fonctionnelle de personnes ayant une acuité auditive normale ou une surdité, ainsi que pour évaluer le rendement de prothèses auditives et d'implants cochléaires (Gifford & Revit, 2010; Hua, Johansson, Jonsson, & Magnusson, 2012; Kuk & Keenan, 2012; Mueller, Weber, & Bellanova, 2011; Oeding & Valente, 2013; Saliba, Nader, El Fata, & Leroux, 2011; Vaillancourt, Laroche, Giguère, Beaulieu, & Legault, 2011; Wolfe et al., 2009). La popularité de ce test est liée à une grande validité apparente puisqu'il met à contribution des stimuli basés sur des phrases, permettant un accès au contexte et à la redondance linguistique, comme dans une situation de conversation normale. Le *HINT* se rapproche ainsi davantage du contexte réel que les tests basés sur des mots. De plus, le *HINT* repose sur une procédure adaptative. Ceci permet d'éviter les effets de plancher et de plafond. Enfin, la version originale anglaise a été traduite dans plus de 13 langues, dont le français canadien (Soli & Wong, 2008; Vaillancourt et al., 2005). Son utilisation dans le cadre d'études multicentriques permet de faciliter la comparaison des résultats entre les langues et les pays.

Or, bien que le *HINT* soit largement utilisé dans plusieurs langues, aucune recension ne regroupe les données métrologiques de ses différentes versions linguistiques afin d'extraire et de comparer ses propriétés métrologiques. Il apparaît important de réaliser cet exercice, qui permettra de positionner chaque version du *HINT* par rapport aux versions disponibles dans d'autres langues. Notamment, cela pourra être utile aux chercheurs et aux professionnels dans l'interprétation des données issues d'essais multicentriques réalisés en plusieurs langues.

Questions de recherche

Cette étude vise donc à extraire et comparer les propriétés métrologiques du *HINT* selon ses différentes versions linguistiques. Les questions suivantes sont explorées: 1- En considérant les données internationales, quelles sont l'applicabilité, la validité, la fidélité et la sensibilité au changement du *HINT* pour mesurer le SRPh, dans le silence et dans le bruit, chez des personnes présentant une acuité auditive normale ou une surdité? 2- Quelles sont les normes disponibles pour les différentes versions du *HINT*?

Méthodologie

Devis

Une revue critique de la littérature traitant des qualités métrologiques du *HINT* a été réalisée en suivant les lignes directrices du protocole de revue systématique proposé par Glasziou (2001). Cet auteur suggère que la réalisation d'une revue systématique doit inclure la rédaction préalable d'un protocole de revue, la formulation de questions de recherche, la réalisation d'une recherche documentaire et d'une collecte de données systématiques et exhaustives, l'évaluation de la qualité scientifique des articles revus et la synthèse des données recueillies. Ces étapes ont été suivies dans le cadre de cette revue. Toutefois, Higgins & Green (2011) précisent qu'une revue systématique doit être complétée par plus d'une personne, ce qui n'est pas le cas de cette revue. La recherche documentaire, la collecte de données, l'évaluation critique des articles et la synthèse des données recueillies ont été réalisées par une seule personne.

Critères d'inclusion et d'exclusion

Les études sélectionnées devaient présenter des données originales issues d'évaluations sur des sujets humains de tout âge, ayant une acuité auditive normale ou une surdité. Dans le cadre de cette revue, l'acuité auditive d'un sujet a été considérée normale lorsque ses seuils auditifs étaient inférieurs à 25 dB HL sur l'ensemble des fréquences audiométriques, soit de 250 à 8 000 Hz. Les sujets ont été considérés atteints de surdité lorsque les seuils auditifs étaient supérieurs ou égaux à 25 dB HL en moyenne aux fréquences 500, 1000 et 2000 Hz et ce, sans égard à l'âge. L'inclusion d'expérimentations effectuées avec des sujets ayant une acuité auditive normale s'avère incontournable, considérant que le développement et la validation du *HINT* original en anglais américain ont été réalisés auprès de cette population (Nilsson et al., 1994).

Pour être sélectionnées, les études devaient traiter spécifiquement du *HINT* et d'au moins l'une des qualités

métrologiques suivantes: l'applicabilité, la validité, la fidélité, la sensibilité au changement et la normalisation. Une définition de ces qualités métrologiques est présentée au Tableau 1. Les textes devaient provenir de périodiques scientifiques revus par des pairs et être rédigés en français ou en anglais. Les chapitres de livres, les actes de colloques et les thèses ont été exclus. Enfin, seules les études ayant utilisé le *HINT* selon la méthode d'administration standard de recherche du SRPh de Nilsson et al. (1994) décrite précédemment ont été retenues.

Stratégie de recherche

Higgins & Green (2011) ainsi que Glasziou (2001) suggèrent aux auteurs de revue systématique de consulter une variété de bases de données afin d'assurer que la recherche documentaire soit la plus exhaustive possible. La consultation des bases de données *Cochrane Library*, *Medline* et *Embase* serait essentielle à toutes les revues systématiques dans le domaine des sciences de la santé (Higgins & Green, 2011). Glasziou (2001) ajoute qu'une consultation complémentaire des bases de données *CINAHL* et *PsycINFO* assurerait une couverture de la majeure partie de la littérature scientifique publiée dans ce domaine.

Dans le cadre de cette étude, les bases de données *Cochrane Library*, *Medline*, *Embase*, *CINAHL* et *PsycINFO* ont donc été interrogées. Les mots-clés suivants ont été utilisés lors de la recherche documentaire: *Hearing in noise test*, *HINT*, *reliability*, *validity*, *responsiveness*, *minimal detectable change*, *minimal clinically important difference*, *factor analysis*, *translation* et *validation*. Ces mots-clés ont été combinés à l'aide d'opérateurs logiques de la manière suivante: [*reliability* OR *validity* OR *responsiveness* OR (*minimal detectable change*) OR (*minimal clinically important difference*) OR (*factor analysis*) OR *translation* OR *validation*] AND [(*hearing in noise test*) OR *HINT*]. Aucune limite de temps n'a été établie.

Procédures et collecte de données

Les titres des études identifiées par la recherche documentaire ont été revus afin d'éliminer les travaux ne correspondant pas aux critères d'inclusion. Les abrégés des articles conservés ont ensuite été révisés pour valider le respect des critères d'inclusion puis les études retenues ont été lues pour en retirer les données liées aux indicateurs métrologiques recherchés. Enfin, une évaluation critique de chaque article a été complétée.

La collecte de données a été réalisée à l'aide de la grille d'extraction de données proposée par Law & MacDermid

Tableau 1. Définition des qualités métrologiques considérées dans cette étude (Fortin, 2010; Laver Fawcett, 2007; Portney & Watkins, 2009).

| Qualité métrologique | Définition |
|-----------------------------|---|
| • Validité | Capacité d'un outil de mesurer ce qu'il est censé mesurer (<i>validity</i>). |
| • Validité apparente | Indique si, subjectivement, l'instrument semble capable de mesurer ce qu'il prétend mesurer et si la méthode qu'il utilise apparaît pertinente (<i>face validity</i>). |
| • Validité de contenu | Capacité de l'outil de bien représenter les différentes dimensions théoriques du phénomène à l'étude (<i>content validity</i>). |
| • Validité de critère | Relation qui existe entre un outil et un autre instrument (<i>mesure étalon</i>) évaluant le même concept (<i>criterion-related validity</i>). |
| • Validité de construit | Étendue avec laquelle un instrument permet d'obtenir des résultats qui sont conformes au construit théorique d'intérêt (<i>construct validity</i>). |
| □ Validité convergente | Capacité de l'outil à obtenir un résultat similaire fortement corrélé à celui obtenu avec un autre outil évaluant une caractéristique semblable (<i>convergent validity</i>). |
| □ Validité divergente | Capacité de l'instrument à obtenir un résultat faiblement ou négativement corrélé à celui obtenu avec un autre outil évaluant une caractéristique différente (<i>divergent validity</i>). |
| □ Validité discriminante | Capacité de l'outil à établir une distinction entre des sujets qui présentent une différence connue (<i>discriminative ou known-groups validity</i>). |
| • Applicabilité | Le contexte et la population pour lesquels l'outil peut être utilisé (<i>applicability</i>). Peut être assimilée à la validité écologique (<i>ecological validity</i>). |
| • Fidélité | Capacité d'un outil de mesure à fournir des valeurs stables, exemptes d'erreur, et à reproduire les mêmes résultats lorsqu'il est utilisé à plusieurs reprises, dans des situations comparables (<i>reliability</i>). Peut être exprimée à l'aide de coefficients de corrélation (fidélité relative) ou en utilisant l'unité de base de la variable étudiée (fidélité absolue). |
| • Fidélité test-retest | Stabilité temporelle de l'outil entre deux ou plusieurs prises de mesure effectuées dans des conditions identiques, auprès des mêmes individus (<i>test-retest reliability</i>). |
| • Fidélité intrajuge | Stabilité des données obtenues avec l'outil par un évaluateur qui observe le même événement à plus d'une reprise (<i>intrarater reliability</i>). |
| • Fidélité interjuges | Degré d'accord qui existe entre les résultats obtenus par plusieurs observateurs qui mesurent le même événement (<i>interrater reliability</i>). |
| • Sensibilité au changement | Capacité d'un outil de mesure à détecter un changement minimal dans la condition d'un patient qui soit cliniquement significatif (<i>responsiveness to change</i>). |

(2008). Cette grille a été développée pour des études portant sur les qualités métrologiques d'outils de mesure. Elle permet de récolter de manière systématique les données au sujet de la population, de l'intervention effectuée au cours de l'expérimentation, de la fidélité relative et absolue, du changement minimal détectable, de la validité de contenu, de critère et de construit, de la cohérence interne, des effets de plancher et de plafond, de la sensibilité au changement, du changement minimal cliniquement important ainsi que de l'applicabilité de l'outil étudié. Lorsque disponible, les données normatives ont également été recueillies.

Évaluation critique des articles

La grille « *Critical appraisal of study design for psychometric articles* » (Law & MacDermid, 2008) a été utilisée pour réaliser l'évaluation critique des articles retenus. Cette grille est constituée de 12 items portant sur la question de recherche, la méthodologie et les procédures utilisées, les analyses effectuées ainsi que les recommandations formulées (voir Tableau 2). Chaque item est coté sur une échelle à trois niveaux, soit 0, 1 ou 2, 2 étant la cote la plus élevée. Le score de chaque item est attribué en suivant les instructions présentées dans le guide accompagnant l'outil, présenté à l'Annexe 1. Le score final est exprimé en pourcentage. Dans le cadre de cette revue, un score de 75% ou plus a été considéré comme un degré de qualité scientifique satisfaisant, à l'instar d'autres auteurs ayant utilisé le même outil (Brudvig, Dirkes, Dutta, & Rane, 2013; Roy, Desmeules, & MacDermid, 2011; Roy, MacDermid, & Woodhouse, 2010).

Résultats

Vingt-quatre articles répondant aux critères d'inclusion ont été examinés. Le processus de recherche documentaire est présenté à la Figure 1. Les articles sélectionnés traitent de 17 versions du *HINT*, dont la version originale anglo-américaine (Duncan & Aarts, 2006; Hanks & Johnson, 1998; Lamothe, Gascon, Larivière, Handfield, & Laroche, 2002; Nilsson et al., 1994; Ribera, 2005; Vermiglio, 2008; Wilson, McArdle, & Smith, 2007), 15 adaptations en diverses langues (Baron de Otero, Brik, Flores, Ortiz, & Abdala, 2008; Bevilacqua, Banhara, Da Costa, Vignoly, & Alvarenga, 2008; Cekic & Sennaroglu, 2008; Hallgren, Larsby, & Arlinger, 2006; Huarte, 2008; Lolov, Raynov, Boteva, & Edrev, 2008; Moon et al., 2008; Myhrum & Moen, 2008; Nielsen & Dau, 2011; Quar et al., 2008; Shiroma, Iwaki, Kubo, & Soli, 2008; Vaillancourt et al., 2005; Wong & Soli, 2005; Wong, Soli, Liu, Han, & Huang, 2007) et la version franco-canadienne pour enfants (Laroche et al., 2006; Vaillancourt, Laroche, Giguere, & Soli, 2008). Un sommaire de ces articles est présenté à l'Annexe 2.

L'évaluation critique montre qu'aucun article n'a obtenu une performance de 75% ou plus, les scores variant entre 45 et 73%. Le sommaire des résultats de cette évaluation est présenté au Tableau 2.

Applicabilité et validité

Deux études ont porté sur l'applicabilité du *HINT* dans des contextes spécifiques (Lamothe et al., 2002; Laroche et al., 2006). Toutes les versions du *HINT* identifiées ont été validées auprès de sujets ayant une acuité auditive normale. Trois études ont évalué la validité du *HINT* auprès de sujets atteints de surdité (Nielsen & Dau, 2011; Vermiglio et al., 2012; Wilson et al., 2007).

Applicabilité. Les 15 adaptations pour lesquelles le test est disponible sont listées au Tableau 3; toutes ces versions ont fait l'objet d'une validation auprès de la population visée.

Une étude a porté sur l'applicabilité du *HINT* anglo-américain auprès d'une population bilingue franco-canadienne (Lamothe et al., 2002). Les résultats suggèrent que cette version du test peut être utilisée avec cette population. Une autre étude traitant d'une version du *HINT* pour enfants franco-canadiens (Laroche et al., 2006) a décelé un effet de l'âge sur les résultats au test, les enfants plus jeunes ayant obtenu un SRPh dans le bruit (condition de face) plus élevé que les enfants plus âgés et les adultes. Les auteurs concluent que des normes spécifiques à l'âge doivent être développées, ce qui a été fait dans une étude subséquente (Vaillancourt et al., 2008).

Validité de contenu. Le *HINT* a été originalement développé en suivant cinq étapes principales (Nilsson et al., 1994): le développement des phrases, la production du bruit de bande large à partir des phrases retenues, l'égalisation de la difficulté des phrases, la création des listes de phrases et l'évaluation de la fidélité. Les différentes versions linguistiques du *HINT* semblent toutes avoir été développées selon un processus cohérent avec le développement du test original. En effet, les études portant sur les versions linguistiques du *HINT* expliquent comment le matériel vocal a été localement développé. Typiquement, les phrases ont été traduites du *HINT* anglo-américain ou extraites de livres pour enfants (Baron de Otero et al., 2008; Cekic & Sennaroglu, 2008; Hallgren et al., 2006; Huarte, 2008; Moon et al., 2008; Myhrum & Moen, 2008; Vaillancourt et al., 2005). Ces phrases ont été révisées par des locuteurs de chaque langue pour assurer leur caractère naturel. Les niveaux de présentation ont été ajustés lors des enregistrements audio afin d'obtenir des scores de reconnaissance de la parole équivalents entre les phrases

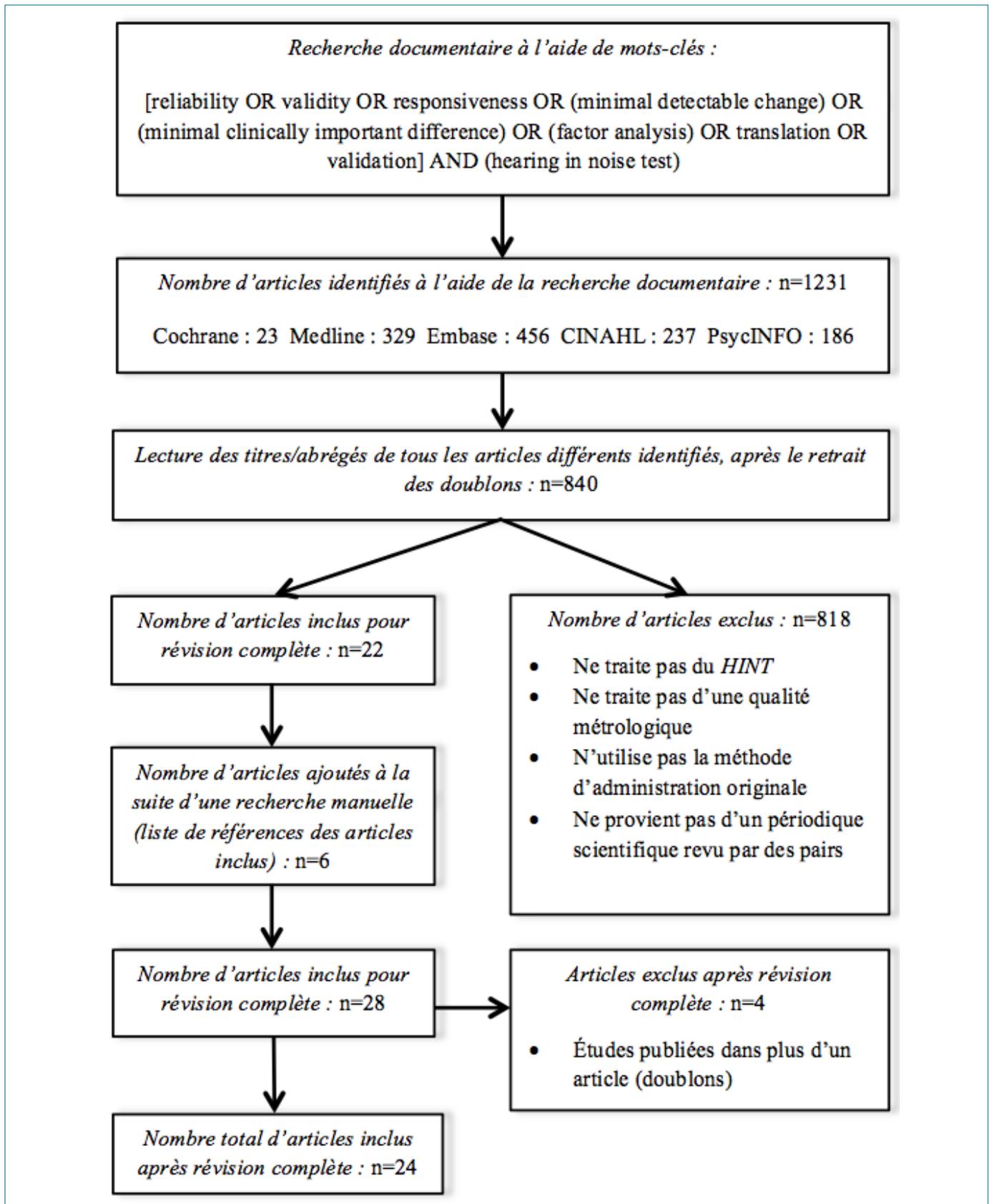


Figure 1. Sommaire de la recherche documentaire

Tableau 2. Sommaire de l'évaluation critique des articles publiés à propos du HINT

| Étude | Score obtenu à chaque item | | | | | | | | | | | | Score total |
|------------------------------|----------------------------|---|---|---|---|-----|---|---|---|----|-----|----|-------------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | |
| Laroche et al. (2006) | 2 | 2 | 1 | 1 | 0 | N/A | 2 | 2 | 2 | 1 | 1 | 2 | 73 |
| Nilsson et al. (1994) | 2 | 1 | 0 | 2 | 1 | N/A | 2 | 2 | 1 | 1 | 2 | 2 | 73 |
| Vaillancourt et al. (2005) | 2 | 1 | 0 | 1 | 1 | N/A | 2 | 2 | 2 | 1 | 2 | 2 | 73 |
| Vaillancourt et al. (2008) | 2 | 2 | 0 | 1 | 1 | N/A | 2 | 2 | 2 | 1 | 1 | 2 | 73 |
| Duncan & Aarts (2006) | 2 | 2 | 0 | 1 | 0 | N/A | 2 | 2 | 2 | 1 | 1 | 2 | 68 |
| Hanks & Johnson (1998) | 2 | 2 | 0 | 1 | 1 | N/A | 2 | 2 | 2 | 1 | 0 | 2 | 68 |
| Cekic & Sennaroglu (2008) | 1 | 2 | 0 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 67 |
| Nielsen & Dau (2011) | 2 | 1 | 0 | 1 | 0 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 67 |
| Wilson et al. (2007) | 2 | 2 | 1 | 0 | 1 | N/A | 2 | 2 | 2 | 0 | 1 | 1 | 64 |
| Wong et al. (2007) | 1 | 2 | 0 | 1 | 1 | N/A | 2 | 1 | 2 | 1 | 1 | 2 | 64 |
| Wong & Soli (2005) | 2 | 1 | 0 | 1 | 1 | N/A | 2 | 1 | 2 | 1 | 1 | 2 | 64 |
| Baron de Otero et al. (2008) | 0 | 1 | 0 | 1 | 1 | N/A | 1 | 2 | 2 | 2 | N/A | 2 | 60 |
| Lamothe et al. (2002) | 2 | 2 | 1 | 1 | 0 | N/A | 2 | 1 | 2 | 1 | 0 | 1 | 59 |
| Myhrum & Moen (2008) | 0 | 2 | 0 | 1 | 1 | N/A | 1 | 1 | 2 | 1 | 1 | 2 | 55 |
| Vermiglio et al. (2012) | 1 | 1 | 1 | 1 | 1 | N/A | 1 | 1 | 1 | 2 | N/A | 1 | 55 |
| Bevilacqua et al. (2008) | 0 | 1 | 0 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 2 | 54 |
| Hallgren et al. (2006) | 1 | 2 | 0 | 1 | 0 | 2 | 1 | 1 | 2 | 1 | 0 | 2 | 54 |
| Huarte (2008) | 1 | 2 | 0 | 1 | 0 | N/A | 1 | 1 | 2 | 1 | 1 | 1 | 50 |
| Lolov et al. (2008) | 0 | 2 | 0 | 1 | 0 | N/A | 1 | 1 | 2 | 1 | 1 | 2 | 50 |
| Moon et al. (2008) | 0 | 2 | 0 | 1 | 1 | N/A | 1 | 1 | 2 | 1 | 1 | 1 | 50 |
| Quar et al. (2008) | 0 | 2 | 0 | 1 | 1 | N/A | 1 | 1 | 2 | 1 | 1 | 1 | 50 |
| Shiroma et al. (2008) | 0 | 2 | 0 | 1 | 1 | N/A | 1 | 1 | 2 | 1 | 1 | 1 | 50 |
| Vermiglio (2008) | 1 | 1 | 0 | 1 | 1 | N/A | 1 | 1 | 2 | 1 | 1 | 1 | 50 |
| Ribera (2005) | 1 | 2 | 0 | 1 | 0 | N/A | 1 | 1 | 2 | 1 | 0 | 1 | 45 |

Note 1. Les items du questionnaire de Law & MacDermid (2008) sont présentés ci-bas (traduction libre):

1. Est-ce que l'article réfère aux recherches scientifiques pertinentes permettant de cerner les connaissances actuelles sur les qualités métrologiques des mesures à l'étude et explique le besoin et les retombées potentielles du projet?
2. Est-ce que les critères d'inclusion et d'exclusion sont pertinents et bien définis?
3. Est-ce que des hypothèses psychométriques spécifiques sont proposées?
4. Est-ce qu'un éventail approprié de qualités métrologiques a été considéré?
5. Est-ce que la taille de l'échantillon est suffisante?
6. Pour les études impliquant un retest ou un suivi, est-ce qu'un taux de suivi / rétention approprié a été obtenu?
7. Est-ce qu'une description de l'outil de mesure, de son application et de son interprétation est fournie? Cette description est-elle suffisamment précise pour permettre la réplication?
8. Est-ce que l'administration et l'application des techniques de mesure dans l'étude ont été standardisées; est-ce que les auteurs ont considéré les sources potentielles d'erreur ou de mauvaise interprétation?
9. Est-ce que des analyses ont été menées pour chaque hypothèse ou objectif de recherche?
10. Est-ce que les tests statistiques réalisés pour obtenir une estimation ponctuelle des qualités métrologiques étaient appropriés?
11. Est-ce que des analyses complémentaires appropriées ont été réalisées pour poursuivre l'analyse au-delà de l'estimation ponctuelle des qualités métrologiques (intervalles de confiance, comparaisons avec des critères, erreur de mesure ou changement minimal détectable)?
12. Est-ce que les conclusions / les recommandations cliniques formulées sont supportées par les objectifs, les analyses et les résultats de l'étude?

Note 2. Lorsqu'un item ne s'appliquait pas à une étude particulière, la mention N/A (non applicable) a été inscrite. Un item non applicable était exclu du calcul du score final, qui était alors calculé en n'utilisant que le nombre d'items restants puis ramené sur 100 %. Ainsi, le nombre d'items total varie d'une étude à l'autre en tenant compte du nombre d'items non-applicables.

(Baron de Otero et al., 2008; Bevilacqua et al., 2008; Cekic & Sennaroglu, 2008; Hallgren et al., 2006; Huarte, 2008; Lolov et al., 2008; Moon et al., 2008; Myhrum & Moen, 2008; Nilsson et al., 1994; Quar et al., 2008; Vaillancourt et al., 2005; Wong & Soli, 2005). Les phrases jugées trop faciles ou difficiles à répéter ont été retirées (Shiroma et al., 2008). Enfin, les phrases ont été groupées pour former des listes finales ayant une distribution phonémique équivalente. Toutes les versions du *HINT* ont été validées auprès de sujets ayant une acuité auditive normale, pour une administration sous écouteurs, sauf la version suédoise (Hallgren et al., 2006) et la version franco-canadienne pour enfants (Laroche et al., 2006) qui ont été validées sous haut-parleurs.

Validité convergente. Deux études ont comparé les résultats de sujets au *HINT* anglo-américain et à l'audiométrie tonale. Les résultats de Wilson et al. (2007) suggèrent une corrélation faible mais significative de 0,323 ($p < 0,01$) entre le SRPh dans le bruit et la moyenne aux sons purs (MSP) pour les fréquences de 500 à 2000 Hz. Vermiglio, Soli, Freed, & Fisher (2012) rapportent pour leur part une corrélation significative de 0,80 ($p < 0,05$) entre le

SRPh dans le silence et la MSP pour les fréquences de 500 à 2000 Hz. Lorsque le calcul de la MSP inclut les hautes fréquences jusqu'à 6000 Hz, ces auteurs mentionnent une corrélation significative variant de 0,63 à 0,77 ($p < 0,05$) pour le SRPh dans le silence ainsi qu'une corrélation faible mais significative variant de 0,37 à 0,46 ($p < 0,05$) pour le SRPh dans le bruit. Vermiglio et al. (2012) ont également comparé les résultats au *HINT* avec l'index articulatoire. Une corrélation significative de 0,66 ($p < 0,05$) est rapportée pour le SRPh dans le silence alors qu'une corrélation plus faible variant de 0,37 à 0,43 ($p < 0,05$) a été obtenue pour le SRPh dans le bruit.

Validité discriminante. Trois études ont vérifié ce type de validité. Dans une étude portant sur le *HINT* anglo-américain (Wilson et al., 2007), les sujets avec surdité ont obtenu un SRPh moyen dans le bruit plus élevé de 5,6 dB par rapport aux sujets normo-entendants; 28% des sujets avec surdité ont toutefois obtenu un score se situant au-delà du 95^{ème} percentile de l'étendue des scores des sujets normo-entendants, ce qui a été considéré par les auteurs comme un résultat normal. Dans une autre étude portant sur la version originale du *HINT* (Vermiglio et al., 2012), une analyse de covariance multivariée a révélé des différences

significatives entre les SRPh moyens dans le silence de sujets normaux et de sujets atteints de surdité variant de 3,59 à 6,65 dB ($p \leq 0,05$) selon le degré de surdité; dans le bruit, seul le groupe de sujets ayant une surdité profonde a obtenu un SRPh moyen (score composite) différent de la normale (+ 1,89 dB, $p \leq 0,05$). Dans l'étude traitant du *HINT* danois (Nielsen & Dau, 2011), aucune différence significative ($p=0,15$) n'a été trouvée entre les SRPh moyens dans le bruit pour les groupes de sujets avec et sans surdité.

Fidélité et sensibilité au changement

Un résumé des données de fidélité est présenté au Tableau 4. Les études ayant évalué la fidélité du *HINT* ont majoritairement été réalisées auprès de sujets ayant une acuité auditive normale, pour une administration sous écouteurs. Deux études ont vérifié la fidélité du *HINT* avec des sujets ayant une surdité (Nielsen & Dau, 2011; Wilson et al., 2007) et deux autres études l'ont fait pour l'administration du test sous haut-parleurs (Hallgren et al., 2006; Laroche et al., 2006). Aucune étude traitant de la sensibilité au changement n'a été recensée.

Fidélité relative. Une étude a évalué la fidélité relative du *HINT* à l'aide de coefficients de corrélation test-retest. Les auteurs du *HINT* japonais (Shiroma et al., 2008) rapportent des coefficients de corrélation de 0,83 pour la présentation du *HINT* sous écouteurs et de 0,92 pour une présentation sous haut-parleurs.

Une seconde étude s'est intéressée à l'accord inter-juges pour l'administration du *HINT* via un service de télé-santé (Ribera, 2005). L'auteur rapporte des coefficients Kappa de 0,96 à 0,98 entre un évaluateur qui était en présence de l'usager et un deuxième qui se trouvait sur un site distant.

Fidélité absolue. L'erreur de mesure standard, parfois accompagnée d'un intervalle de confiance à 95%, est l'indicateur de fidélité absolue le plus souvent rapporté. Pour la condition d'écoute dans le silence, l'erreur de mesure varie de 0,79 à 3,1 dB avec un intervalle de confiance de 1,54 à 6,0 dB. Pour les conditions d'écoute dans le bruit, elle varie de 0,43 à 1,9 dB avec un intervalle de confiance de 0,8 à 3,7 dB. Une étude sur le *HINT* anglo-américain rapporte un changement minimal détectable de 2,8 dB pour la condition de présentation dans le silence et de 2,2 dB pour les conditions de bruit (Nilsson et al., 1994).

Un second indicateur de fidélité absolue rapporté est la fidélité inter-listes, soit la capacité du test à produire un SRPh équivalent entre deux passations chez un même sujet lorsque deux listes de phrases différentes sont utilisées. Cet

Tableau 3. Adaptations du *HINT* anglo-américain

| Langues | Auteurs |
|---------------------------|------------------------------|
| Bulgare | Lolov et al. (2008) |
| Cantonais | Wong & Soli (2005) |
| Coréen | Moon et al. (2008) |
| Danois | Nielsen & Dau (2011) |
| Espagnol castillan | Huarte (2008) |
| Espagnol latino-américain | Baron de Otero et al. (2008) |
| Français canadien | Vaillancourt et al. (2005) |
| Japonais | Shiroma et al. (2008) |
| Malais | Quar et al. (2008) |
| Mandarin continental | Wong et al. (2007) |
| Mandarin taïwanais | Wong et al. (2007) |
| Norvégien | Myhrum & Moen (2008) |
| Portugais brésilien | Bevilacqua et al. (2008) |
| Suédois | Hallgren et al. (2006) |
| Turc | Cekic & Sennaroglu (2008) |

indicateur est important parce que l'administration du *HINT* en clinique se réalise habituellement en n'utilisant qu'une partie des listes de phrases disponibles. Certains auteurs ont voulu déterminer si un SRPh mesuré à l'aide d'une liste de phrases donnée serait identique à ceux obtenus à l'aide d'autres listes. Les résultats des études ayant évalué cet aspect (Hallgren et al., 2006; Hanks & Johnson, 1998; Laroche et al., 2006; Nielsen & Dau, 2011; Nilsson et al., 1994; Quar et al., 2008; Vaillancourt et al., 2005; Wong & Soli, 2005) suggèrent que les différentes listes de phrases produisent des SRPh semblables.

Normalisation

Les normes disponibles pour chaque version du *HINT* sont présentées au Tableau 5. Pour une administration sous écouteurs auprès de sujets normo-entendants, le SRPh moyen dans le silence varie de 14,6 à 25,9 dB(A) selon la version du test, avec un écart-type de 2,2 à 7,5 dB et une moyenne de 18,4 dB(A). Dans le bruit (score composite), il

Tableau 4. Fidélité – Résumé des données recueillies

| Indicateur métrologique | Données extraites |
|-------------------------|---|
| Fidélité relative | Coefficients de corrélation test-retest: - 0,92 (sous haut-parleurs) (Shiroma et al., 2008) - 0,83 (sous écouteurs) (Shiroma et al., 2008) Coefficients Kappa: - 0,96 à 0,98 (Ribera, 2005) |
| Fidélité absolue | Erreur de mesure standard (SEM) et intervalles de confiance à 95% (CI) - SEM ± CI (listes de 20 phrases): |
| | Dans le silence: |
| | - 0,79 dB ± 1,54 (Wong et al., 2007) |
| | - 0,82 dB ± 1,62 (Wong et al., 2007) |
| | - 0,99 dB ± 1,94 (Nilsson et al., 1994) |
| | - 1,1 dB ± 2,1 (Lolov et al., 2008) |
| | - 1,2 dB (Myhrum & Moen, 2008) |
| | - 1,5 dB ± 2,9 (Bevilacqua et al., 2008) |
| | - 1,6 dB ± 3,1 (Cekic & Sennaroglu, 2008) |
| | - 1,7 dB ± 2,3 (Wong & Soli, 2005) |
| | - 2,0 dB ± 3,3 (Quar et al., 2008) |
| | - 2,0 dB ± 3,9 (Vermiglio, 2008) |
| | - 2,2 dB (Vaillancourt et al., 2005) |
| | - 3,1 dB ± 6,0 (Moon et al., 2008) |
| | Dans le bruit: |
| | - 0,43 ± 0,80 (Duncan & Aarts, 2006) |
| | - 0,6 dB (Vaillancourt et al., 2005) |
| | - 0,6 dB ± 1,2 (Lolov et al., 2008) |
| | - 0,68 dB (Hallgren et al., 2006) |
| | - 0,7 dB (Vaillancourt et al., 2008) |
| | - 0,75 dB ± 1,47 (Wong et al., 2007) |
| | - 0,76dB ± 1,49 (Nilsson et al., 1994) |
| | - 0,86 dB (NF) (Nielsen & Dau, 2011) |
| | - 0,89 dB ± 1,74 (Wong et al., 2007) |
| | - 0,9 dB (Myhrum & Moen, 2008) |
| | - 0,92 dB pour adultes avec surdit  (NF) (Nielsen & Dau, 2011) |
| | - 1,0 dB ± 2,0 (Cekic & Sennaroglu, 2008) |

| |
|--|
| - 1,1 dB pour adultes (Laroche et al., 2006) |
| - 1,1 dB \pm 1,6 (Wong & Soli, 2005) |
| - 1,2 dB pour enfants (Laroche et al., 2006) |
| - 1,2 dB \pm 2,3 (Bevilacqua et al., 2008) |
| - 1,3 dB \pm 2,1 (Quar et al., 2008) |
| - 1,4 dB après 1 semaine (Hallgren et al., 2006) |
| - 1,5 \pm 3,0 (Vermiglio, 2008) |
| - 1,9 dB \pm 3,7 (Moon et al., 2008) |
| Changement minimal détectable (MDC): |
| - 2,8 dB (silence) (Nilsson et al., 1994) |
| Équivalence des listes de phrases: |
| - Selon ANOVA, pas de différence entre les listes ($p > 0,05$) (Hallgren et al., 2006; Hanks & Johnson, 1998; Laroche et al., 2006; Nielsen & Dau, 2011; Nilsson et al., 1994; Quar et al., 2008; Vaillancourt et al., 2005) |
| - Écart entre SRPh moyen de chaque liste et de toutes les listes $< \pm 1$ dB (Hallgren et al., 2006; Laroche et al., 2006; Vaillancourt et al., 2005; Wong & Soli, 2005) |

varie de -8,8 à -6,4 dB(S/B), avec un écart-type de 0,57 à 2,5 dB et une moyenne de -7,6 dB(S/B). Ces SRPh moyens \pm 2 écarts-types, afin d'estimer un intervalle de confiance à 95 %, sont également présentés aux Figures 2 et 3.

Discussion

Le *HINT* original a été adapté en 15 langues. Il est plutôt rare dans le domaine de l'audiométrie vocale qu'un test servant à mesurer les capacités de reconnaissance de la parole dans le bruit soit rendu disponible en autant de versions linguistiques différentes. Il convient ainsi de souligner les efforts considérables qui ont dû être déployés par les équipes de recherche ayant travaillé à l'adaptation linguistique du *HINT* de par le monde.

Plusieurs articles expliquent en détail le développement des différentes versions du test, mettant en évidence que les différentes versions du test ont toutes été développées d'une manière systématique, en suivant un processus similaire. La majorité des versions recensées ont fait l'objet d'études de fidélité et de normalisation; ainsi, l'erreur de mesure et les normes du test sont bien connues. Le *HINT* original en anglais américain pourrait par ailleurs être utilisé auprès de sujets francophones bilingues et sa version franco-canadienne chez des enfants âgés de 6 ans ou plus. Ces qualités font du *HINT* un outil attrayant,

particulièrement pour une utilisation dans le contexte d'études multicentriques.

Toutefois, les scores de qualité attribués aux articles lors de l'évaluation critique sont tous inférieurs au critère de qualité préétabli (75 %). Plusieurs éléments peuvent expliquer ce constat, dont notamment l'utilisation d'approches similaires pour démontrer les propriétés métrologiques du test et des effectifs limités, particulièrement pour la vérification de la fidélité et l'établissement de normes.

Selon Portney & Watkins (2009), la validation d'un outil de mesure est un processus qui requiert l'utilisation de diverses méthodes. Les validités apparente et de contenu d'un outil sont habituellement considérées lors de son développement. Ces paramètres sont toutefois insuffisants et doivent être complétés par une étude de la validité de critère ou de construit. La validité de contenu de toutes les versions du *HINT* a été bien démontrée. Par contre, seules les versions anglo-américaine (Nilsson et al., 1994; Vermiglio et al., 2012; Wilson et al., 2007) et danoise (Nielsen & Dau, 2011) du *HINT* ont fait l'objet d'une étude de la validité de construit.

Law & MacDermid (Law & MacDermid, 2008) mentionnent par ailleurs que l'évaluation métrologique d'un

Tableau 5. Moyenne et écart-type obtenus au HINT dans les quatre conditions d'écoute et pour le score composite dans diverses langues

| Langue | Silence (dBA) | | Bruit face (dB S/B) | | Bruit droite (dB S/B) | | Bruit gauche (dB S/B) | | Composite (dB S/B) | |
|---|---------------|------|---------------------|------|-----------------------|------|-----------------------|------|--------------------|------|
| | M | ET | M | ET | M | ET | M | ET | M | ET |
| Anglais américain (Vermiglio, 2008) | 15.6 | 3.1 | -2.6 | 1.0 | -10.1 | 1.3 | -10.1 | 1.3 | -6.4 | 0.9 |
| Bulgare (Lolov et al., 2008) | 24.2 | 7.5 | -4.0 | 1.5 | -9.7 | 3.2 | -10.2 | 3.9 | -7.0 | 2.5 |
| Cantonais (Wong & Soli, 2005) | 19.4 | 3.1 | -4.0 | 0.9 | -10.9 | 1.0 | -11.0 | 0.8 | -7.5 | 0.7 |
| Castillan (Huarte, 2008) | 14.6 | 3.2 | -3.6 | 1.2 | -11.8 | 1.2 | -12.1 | 1.0 | -7.7 | 0.8 |
| Coréen (Moon et al., 2008) | 18.6 | 3.0 | -3.3 | 1.0 | -10.8 | 1.1 | -10.5 | 1.2 | -7.0 | 0.7 |
| Danois (Nielsen & Dau, 2011) | – | – | -2.52 | 0.87 | – | – | – | – | – | – |
| Français canadien (Vaillancourt et al., 2005) | 16.4 | 3.8 | -3.0 | 1.1 | -11.4 | 1.3 | -11.4 | 1.3 | -7.2 | 0.8 |
| Japonais (Shiroma et al., 2008) | 16.8 | 3.1 | -5.3 | 1.4 | -12.2 | 1.4 | -12.5 | 1.4 | -8.8 | 1.1 |
| Malais (Quar et al., 2008) | 20.2 | 3.4 | -4.7 | 0.8 | -12.4 | 1.0 | -12.4 | 1.0 | -8.6 | 0.7 |
| Mandarin continental (Wong et al., 2007) | 14.7 | 1.49 | -4.3 | 0.62 | -11.7 | 0.85 | -11.7 | 0.79 | -8.0 | 0.57 |
| Mandarin taiwanais (Wong et al., 2007) | 19.4 | 3.1 | -4.0 | 0.94 | -10.9 | 1.0 | -11.0 | 0.8 | -7.5 | 0.72 |
| Norvégien (Myhrum & Moen, 2008) | 17.5 | 2.9 | -3.2 | 1.0 | -10.3 | 1.1 | -10.4 | 1.1 | -6.7 | 0.8 |
| Portugais brésilien (Bevilacqua et al., 2008) | 15.3 | 3.5 | -4.6 | 0.8 | -12.1 | 1.0 | -12.2 | 0.8 | -8.4 | 0.6 |
| Turc (Cekic & Sennaroglu, 2008) | 25.9 | 2.2 | -3.9 | 0.9 | -11.8 | 1.1 | -12.1 | 1.0 | -7.9 | 0.8 |
| Étendue: | 14.6 | 2.2 | -5.3 | 0.62 | -12.4 | 0.85 | -12.5 | 0.79 | -8.8 | 0.57 |
| | à | à | à | à | à | à | à | à | à | à |
| | 25.9 | 7.5 | -2.52 | 1.5 | -9.7 | 3.2 | -10.1 | 3.9 | -6.4 | 2.5 |

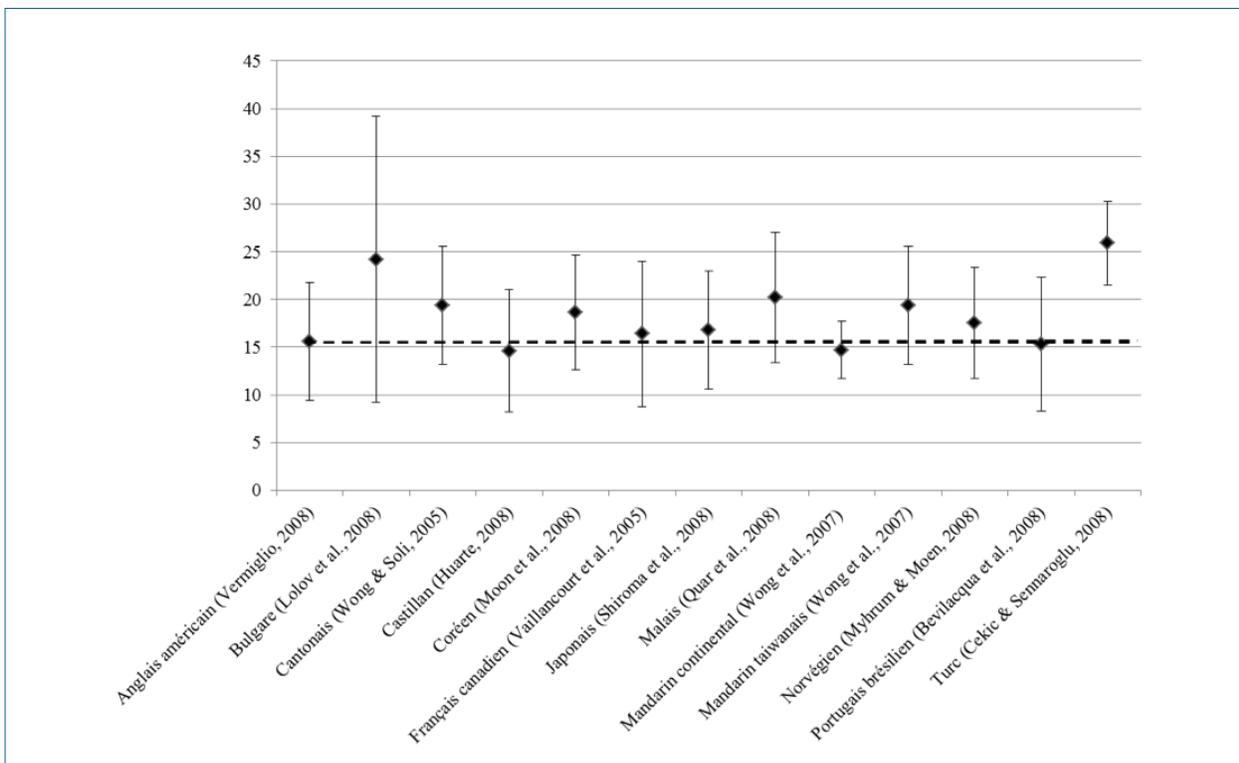


Figure 2. Normes (± 2 écarts-types) du *HINT* obtenues dans le silence, sous écouteurs (dBA) dans diverses langues. La ligne pointillée représente le SRPh moyen dans le silence en anglais américain.

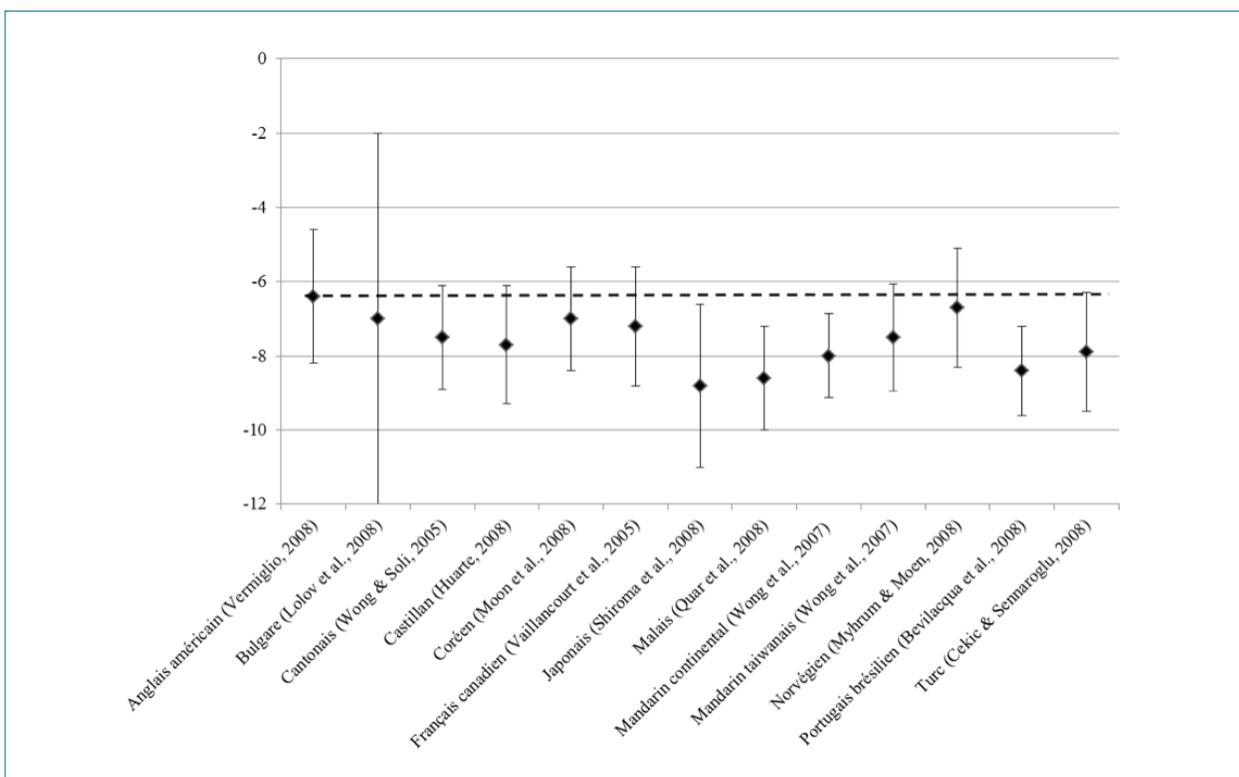


Figure 3. Normes (± 2 écarts-types) relatives au score composite du *HINT*, sous écouteurs (dB S/B) dans diverses langues. La ligne pointillée représente le SRPh moyen dans le bruit (score composite) en anglais américain.

outil de mesure devrait s'intéresser à plus d'une forme de fidélité (test-retest, intra ou inter-évaluateur) et examiner tant la fidélité absolue que la fidélité relative en utilisant plus d'un indicateur (coefficient de corrélation intra-classe, erreur de mesure et changement minimal détectable). Or, la majorité des études ne rapportent que l'erreur de mesure, accompagnée ou non d'un intervalle de confiance à 95%. Une seule étude rapporte un coefficient de corrélation test-retest (Shiroma et al., 2008) mais les auteurs n'ont pas spécifié l'erreur de mesure. Le changement minimal détectable n'est présenté que dans l'article traitant du *HINT* anglo-américain original (Nilsson et al., 1994); la sensibilité du *HINT* au changement ne semble pas avoir été étudiée. À l'instar de l'évaluation de la validité, la vérification de la fidélité et de la sensibilité au changement des différentes versions du *HINT* semble donc présenter des limites.

Un autre facteur ayant diminué la qualité des études revues est le nombre limité de sujets lors de la vérification de la fidélité et de l'établissement de normes pour le test. Bien qu'il soit courant de recruter entre 30 et 50 sujets pour vérifier la fidélité ou développer les normes d'un outil de mesure, il ne semble pas y avoir de consensus dans la communauté scientifique sur le nombre optimal de sujets à inclure (Streiner & Norman, 2008). D'une part, certains auteurs suggèrent d'utiliser le plus grand nombre de sujets possible; Charter (1999, 2003) par exemple propose un minimum de 400 participants. Par opposition, Cicchetti (1999, 2001) avance que 50 participants est un nombre suffisant du point de vue statistique pour la plupart des outils de mesure puisque selon lui, les bénéfices apportés par l'inclusion d'un plus grand nombre de participants seraient minimales en rapport avec les coûts supplémentaires engendrés. Une autre alternative serait d'estimer, pour chaque étude, le nombre de sujets nécessaires à l'aide d'un calcul de taille d'échantillon, basé, entre autres, sur une estimation a priori de la valeur attendue de l'indicateur de fidélité sélectionné (Streiner & Norman, 2008). Ces auteurs proposent de considérer cette dernière option lors de la détermination du nombre de sujets de l'étude, tout en reconnaissant que le critère de 50 participants proposé par Cicchetti (1999, 2001) est probablement suffisant dans la majorité des cas. Parmi les 22 études ayant évalué la fidélité du *HINT* ou proposant des normes, seulement quatre avaient un effectif de 50 sujets ou plus (Myhrum & Moen, 2008; Shiroma et al., 2008; Vermiglio, 2008; Wong et al., 2007); six études ont recruté entre 30 et 49 participants (Cekic & Sennaroglu, 2008; Moon et al., 2008; Quar et al., 2008; Vaillancourt et al., 2005; Wong & Soli, 2005; Wong et al., 2007) et 12 ont été réalisées avec moins de 30 sujets (Bevilacqua et al., 2008; Duncan & Aarts, 2006; Hallgren et al., 2006; Hanks

& Johnson, 1998; Huarte, 2008; Laroche et al., 2006; Lolov et al., 2008; Nielsen & Dau, 2011; Nilsson et al., 1994; Ribera, 2005; Vaillancourt et al., 2008; Wilson et al., 2007). Peu d'études présentaient une justification ou un calcul de la taille d'échantillon.

En ce qui a trait à l'erreur de mesure, cet indicateur semble élevé pour certaines versions du *HINT*. Brand & Kollmeier (2002) suggèrent que l'erreur de mesure, pour des outils d'audiométrie vocale mesurant le SRPh dans le bruit, devrait être inférieure à 1 dB, pour que le test soit en mesure de détecter des différences de performance subtiles, notamment entre des conditions d'écoute différentes ou divers degrés de surdité. Selon Theunissen, Swanepoel & Hanekom (2009), cette valeur est perçue comme un critère de variabilité acceptable et réaliste pour ce type d'outil, plusieurs auteurs ayant rapporté une erreur de mesure de cette envergure. Les travaux de Plomp & Mimpfen (1979) suggèrent par ailleurs que ce critère pourrait aussi être applicable pour la mesure du SRPh dans le silence. Sur cette base, l'erreur de mesure de plusieurs versions du *HINT* apparaît donc élevée. En fait, seule la fidélité des versions du *HINT* en anglais américain (Nilsson et al., 1994), en mandarin taïwanais et en mandarin continental (Wong et al., 2007) serait acceptable. La fidélité des versions du *HINT* en français canadien (Vaillancourt et al., 2005), en bulgare (Lolov et al., 2008), en suédois (Hallgren et al., 2006), en danois (Nielsen & Dau, 2011), en norvégien (Myhrum & Moen, 2008) et en turc (Cekic & Sennaroglu, 2008) serait elle aussi acceptable, mais pour une administration dans les conditions de bruit uniquement.

Par ailleurs, une certaine variabilité est observée dans les normes sous écouteurs rapportées par les auteurs des différentes versions du *HINT*. En considérant les intervalles de confiance à 95 % de chaque version du test et en utilisant comme standard de référence la moyenne de la version anglo-américaine du *HINT* (Vermiglio, 2008), le test ayant été initialement développé dans cette langue, la version turque se démarque pour la condition d'écoute dans le silence. En effet, l'intervalle de confiance de ce test n'englobe pas la valeur de référence (voir Figure 2). Dans le bruit (voir Figure 3), les versions en japonais, malais, mandarin continental et portugais apparaissent différentes de la version originale. Le *HINT* bulgare (Lolov et al., 2008) présente l'intervalle de confiance le plus large en raison de sa variance élevée autant dans le silence que dans le bruit ($ET \pm 7,5$ dB dans le silence et de $\pm 2,5$ dB dans le bruit); l'intervalle à 95 % des normes bulgares est tellement large qu'il englobe les normes de toutes les autres versions du test. En somme, les normes

de certaines versions linguistiques du *HINT* ne sont pas comparables au standard anglophone.

Les différences dans les données normatives pourraient s'expliquer par des niveaux d'acuité auditive ou d'âge des participants variables selon les études (Soli & Wong, 2008; Vermiglio et al., 2012). En effet, il est possible que des individus considérés normo-entendants mais présentant de moins bons seuils auditifs (par exemple, entre 20 et 25 dB HL) puissent obtenir des SRPh plus élevés, particulièrement dans le silence. Par ailleurs, des individus trop âgés pourraient présenter une perte auditive en hautes fréquences pouvant passer inaperçue lors des tests de dépistage audiométrique réalisés lors du recrutement, ou avoir des difficultés de traitement auditif liés au vieillissement; ces deux facteurs pourraient élever les SRPh dans le bruit. Or, il n'y a pas suffisamment d'information sur ces variables dans la majorité des articles revus pour permettre de vérifier cette hypothèse. La plupart des auteurs mentionnent le critère de seuil auditif maximal ayant été accepté chez leurs sujets, mais aucun ne présente l'audiogramme moyen des participants. Quant à l'âge des sujets, les auteurs rapportent généralement l'étendue d'âge de leurs participants, mais seulement quelques auteurs spécifient l'âge moyen. Selon les données disponibles, il est tout de même possible de constater que les échantillons utilisés par les auteurs des versions du *HINT* en japonais et en mandarin continental semblent plus jeunes que les sujets de Vermiglio et al. (2008) (en moyenne 22,5 ans pour le *HINT* japonais et 22 ans pour le *HINT* en mandarin continental, par rapport à 38,7 ans pour le *HINT* anglo-américain), ce qui pourrait expliquer les meilleures performances dans le bruit des deux premiers groupes. Néanmoins, il est difficile de juger de l'homogénéité de l'échantillonnage entre la majorité des études. Par ailleurs, Lolov et al. (2008) n'ont recruté que 18 sujets pour établir les normes bulgares, ce qui pourrait expliquer la grande variance observée.

La variabilité entre les données normatives des différentes versions du *HINT* a aussi été rapportée par Soli & Wong (2008). Ces derniers proposent d'utiliser un score standardisé, le H-score, pour permettre une comparaison directe des résultats de différentes versions du *HINT* et ainsi contrôler la variabilité due à la langue. Toutefois, tout en reconnaissant les écarts entre les versions, Soli & Wong (2008) jugent que leurs normes restent similaires. Les résultats de cette revue ne permettent pas de soutenir cette affirmation.

Enfin, pratiquement toutes les études de validité et de fidélité sur les différentes versions du *HINT* ont été réalisées auprès de sujets ayant une acuité auditive

normale. Seulement trois études recensées ont évalué ces propriétés chez des sujets ayant une surdité (Nielsen & Dau, 2011; Vermiglio et al., 2012; Wilson et al., 2007). Or, le *HINT* est principalement utilisé en clinique auprès de cette population, notamment pour vérifier les capacités de reconnaissance de la parole ou pour mesurer le rendement d'aides auditives. Bien que Nielsen & Dau (2011) rapportent une erreur de mesure très similaire entre leurs sujets avec ou sans surdité (0,92 dB vs 0,86 dB respectivement) et que les résultats de Nilsson et al. (1994) suggèrent que la fidélité du *HINT* pourrait être assez robuste même pour des sujets atteints de surdité, il n'est pas exclu qu'une variabilité différente puisse être observée chez les personnes présentant une surdité. En fait, peu de données sont disponibles à propos des qualités métrologiques du *HINT* lorsqu'utilisé auprès de ces personnes, en particulier pour une administration sous haut-parleurs qui est la procédure courante pour l'évaluation des aides auditives. Ce manque de données métrologiques, notamment en ce qui a trait à la fidélité qui pourrait être affectée par une variabilité différente, constitue une limite majeure à l'utilisation du test auprès de cette population. Dans ce contexte, la fidélité du *HINT* ainsi que sa sensibilité au changement devraient être davantage étudiées.

Grille d'évaluation des qualités métrologiques

Bien que la grille d'évaluation de Law & MacDermid (2008) apparaisse pertinente dans le contexte de cette revue, le processus de cotation semble sévère, en particulier pour l'Item 5 relatif à la taille de l'échantillon. En effet, pour qu'une étude obtienne la cote maximale de 2, Law & MacDermid (2008) demandent que les auteurs aient présenté un calcul de taille d'échantillon, qu'ils aient réussi à recruter le nombre de sujets calculé et qu'ils aient ensuite réalisé une analyse post-hoc de puissance observée. Une étude obtiendra la cote de 1 si les auteurs ont présenté une justification de la taille d'échantillon, sans avoir réalisé de calcul formel de taille d'échantillon ou de puissance observée, ou s'ils ont eu recours à plus de 100 participants, sans en présenter de justification. La cote 0 sera attribuée à une étude dans laquelle les auteurs ne présentent aucune justification de la taille d'échantillon ou lorsque l'étude manque de manière évidente de puissance statistique. À la lumière de la discussion présentée précédemment à propos du nombre de participants à inclure qui soit suffisant du point de vue statistique lors de la vérification des qualités métrologiques d'outils de mesure, il serait préférable que la cote de 1 puisse être attribuée dès que la taille de l'échantillon est égale ou supérieure à 50 sujets (au lieu de 100), ce qui permettrait une évaluation plus réaliste de la qualité scientifique des articles revus.

Limites de la présente étude

Cette étude a porté sur les propriétés métrologiques des différentes versions linguistiques du HINT telles qu'elles sont utilisées en clinique et en recherche. Ainsi, la qualité de l'adaptation linguistique réalisée n'a pas été évaluée à proprement parler, au-delà du constat que la méthode d'adaptation semble uniforme d'une version à l'autre et conforme au processus de développement initial du test. À cet égard, il serait éventuellement pertinent d'évaluer la valeur de la méthode usuelle d'adaptation linguistique du HINT afin de correctement guider de futures adaptations du test.

Par ailleurs, la recherche documentaire, la collecte de données, l'évaluation critique des articles et la synthèse des données recueillies, dans le cadre de cette étude, ont été réalisées par une seule personne. Selon le standard méthodologique des revues systématiques, ces étapes doivent être effectuées par deux évaluateurs indépendants. L'évaluateur unique possédait toutefois de l'expérience pour ce type d'analyse.

Conclusion

Les faits scientifiques disponibles suggèrent que les propriétés métrologiques des différentes versions du HINT ne sont pas toutes bien établies. L'applicabilité, la validité, la fidélité et la sensibilité au changement du HINT demeurent notamment méconnues chez les personnes présentant une surdité. Il s'agit de lacunes importantes qui doivent faire l'objet de nouvelles recherches, considérant que de nombreux cliniciens et chercheurs utilisent le HINT sur une base régulière auprès de sujets ayant une surdité. Par ailleurs, dans un contexte où les versions linguistiques du HINT ne semblent pas toutes présenter des normes directement comparables ni le même niveau de qualité, il est important que les comparaisons entre les résultats obtenus avec différentes versions du HINT soient réalisées de manière prudente, en considérant les propriétés métrologiques spécifiques à chaque version utilisée.

Références

- Baron de Otero, C., Brik, G., Flores, L., Ortiz, S., & Abdala, C. (2008). The Latin American Spanish hearing in noise test. *International Journal of Audiology, 47*, 362-363. doi: 10.1080/14992020802060888
- Bess, F. H., & Humes, L. (2008). *Audiology: The fundamentals* (4th ed.). Philadelphia, PA: Lippincott Williams & Wilkins.
- Bevilacqua, M. C., Banhara, M. R., Da Costa, E. A., Vignoly, A. B., & Alvarenga, K. F. (2008). The Brazilian Portuguese hearing in noise test. *International Journal of Audiology, 47*, 364-365. doi: 10.1080/14992020701870205
- Brand, T., & Kollmeier, B. (2002). Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. *Journal of the Acoustical Society of America, 111*, 2801-2810.
- Brudvig, T. J., Dirkes, A., Dutta, P., & Rane, K. (2013). Critical thinking skills in health care professional students: A systematic review. *Journal of Physical Therapy Education, 27*(3), 12-25.
- Cekic, S., & Sennaroglu, G. (2008). The Turkish hearing in noise test. *International Journal of Audiology, 47*, 366-368. doi: 10.1080/14992020802056027
- Charter, R. A. (1999). Sample size requirements for precise estimates of reliability, generalizability, and validity coefficients. *Journal of Clinical and Experimental Neuropsychology, 21*, 559-566. doi: 10.1076/jcen.21.4.559.889
- Charter, R. A. (2003). Study samples are too small to produce sufficiently precise reliability coefficients. *Journal of General Psychology, 130*, 117-129.
- Cicchetti, D. V. (1999). Sample size requirements for increasing the precision of reliability estimates: Problems and proposed solutions. *Journal of Clinical and Experimental Neuropsychology, 21*, 567-570. doi: 10.1076/jcen.21.4.567.886
- Cicchetti, D. V. (2001). Methodological commentary. The precision of reliability and validity estimates re-visited: Distinguishing between clinical and statistical significance of sample size requirements. *Journal of Clinical and Experimental Neuropsychology, 23*, 695-700. doi: 10.1076/jcen.23.5.695.1249
- Duncan, K. R., & Aarts, N. L. (2006). A comparison of the HINT and Quick SIN tests. *Journal of Speech-Language Pathology & Audiology, 30*, 86-94.
- Fortin, M. F. (2010). *Fondements et étapes du processus de recherche: Méthodes quantitatives et qualitatives* (2^{ème} éd.). Montréal, Canada: Chenelière Éducation.
- Gifford, R. H., & Revit, L. J. (2010). Speech perception for adult cochlear implant recipients in a realistic background noise: Effectiveness of preprocessing strategies and external options for improving speech recognition in noise. *Journal of the American Academy of Audiology, 21*, 441-451. doi: 10.3766/jaaa.21.7.3
- Glasziou, P. (2001). *Systematic reviews in health care: A practical guide*. Cambridge, UK: Cambridge University Press.
- Hallgren, M., Larsby, B., & Arlinger, S. (2006). A Swedish version of the hearing in noise test (HINT) for measurement of speech recognition. *International Journal of Audiology, 45*, 227-237. doi: 10.1080/14992020500429583
- Hanks, W. D., & Johnson, G. D. (1998). HINT list equivalency using older listeners. *Journal of Speech, Language & Hearing Research, 41*, 1335-1340.
- Higgins, J. P. T., & Green, S. (Eds.). (2011). *Cochrane handbook for systematic reviews of interventions version 5.1.0*. <http://handbook.cochrane.org>
- Hua, H., Johansson, B., Jonsson, R., & Magnusson, L. (2012). Cochlear implant combined with a linear frequency transposing hearing aid. *Journal of the American Academy of Audiology, 23*, 722-732. doi: 10.3766/jaaa.23.9.6
- Huarte, A. (2008). The Castilian Spanish hearing in noise test. *International Journal of Audiology, 47*, 369-370. doi: 10.1080/14992020801908269

- Kuk, F., & Keenan, D. (2012). Efficacy of a reverse cardioid directional microphone. *Journal of the American Academy of Audiology*, 23, 64-73. doi: 10.3766/jaaa.23.1.7
- Lamothe, J., Gascon, C., Larivière, M., Handfield, M., & Laroche, C. (2002). Standardisation of the hearing in noise test (HINT) for a bilingual Francophone population and an Anglophone population. *Journal of Speech-Language Pathology & Audiology*, 26, 81-89.
- Laroche, C., Vaillancourt, V., Melanson, C., Renault, M. E., Thériault, C., Soli, S. D., & Giguère, C. (2006). Adaptation du HINT (hearing in noise test) pour les enfants francophones canadiens et données préliminaires sur l'effet d'âge. *Journal of Speech-Language Pathology and Audiology*, 30, 95-109.
- Laver Fawcett, A. J. (2007). *Principles of assessment and outcome measurement for occupational therapists and physiotherapists: Theory, skills and application*. John Wiley & Sons Ltd, Chichester, England.
- Law, M., & MacDermid, J. (2008). *Evidence-based rehabilitation: A guide to practice* (2nd ed.). Thorofare, NJ: Slack Incorporated.
- Lolov, S. R., Raynov, A. M., Boteva, I. B., & Edrev, G. E. (2008). The Bulgarian hearing in noise test. *International Journal of Audiology*, 47, 371-372. doi: 10.1080/14992020801886788
- McArdle, R., & Hnath-Chilsum, T. (2009). Speech audiometry. In J. Katz, R. Burkard, L. Hood & L. Medwetsky (Eds.), *Handbook of clinical audiology* (6th ed., pp. 64-79). Philadelphia, PA: Lippincott Williams & Wilkins.
- Moon, S. K., Hee Kim, S., Ah Mun, H., Jung, H. K., Lee, J. H., Choung, Y. H., & Park, K. (2008). The Korean hearing in noise test. *International Journal of Audiology*, 47, 375-376. doi: 10.1080/14992020701882457
- Mueller, H. G., Weber, J., & Bellanova, M. (2011). Clinical evaluation of a new hearing aid anti-cardioid directivity pattern. *International Journal of Audiology*, 50, 249-254. doi: 10.3109/14992027.2010.547992
- Myhrum, M., & Moen, I. (2008). The Norwegian hearing in noise test. *International Journal of Audiology*, 47, 377-378. doi: 10.1080/14992020701876707
- Nielsen, J. B., & Dau, T. (2011). The Danish hearing in noise test. *International Journal of Audiology*, 50, 202-208. doi: 10.3109/14992027.2010.524254
- Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. *Journal of the Acoustical Society of America*, 95, 1085-1099.
- Oeding, K., & Valente, M. (2013). Sentence recognition in noise and perceived benefit of noise reduction on the receiver and transmitter sides of a BICROS hearing aid. *Journal of the American Academy of Audiology*, 24, 980-991. doi: 10.3766/jaaa.24.10.9
- Plomp, R., & Mimpen, A. M. (1979). Improving the reliability of testing the speech reception threshold for sentences. *Audiology*, 18, 43-52.
- Portney, L. G., & Watkins, M. P. (2009). *Foundations of clinical research: Applications to practice* (3rd ed.). Upper Saddle River, NJ: Pearson Education.
- Quar, T. K., Mukari, S. Z., Abdul Wahab, N. A., Abdul Razak, R., Omar, M., & Maamor, N. (2008). The Malay hearing in noise test. *International Journal of Audiology*, 47, 379-380. doi: 10.1080/14992020801886796
- Ribera, J. E. (2005). Interjudge reliability and validation of telehealth applications of the hearing in noise test. *Seminars in Hearing*, 26, 13-18.
- Roy, J. S., Desmeules, F., & MacDermid, J. C. (2011). Psychometric properties of presenteeism scales for musculoskeletal disorders: A systematic review. *Journal of Rehabilitation Medicine*, 43, 23-31. doi: 10.2340/16501977-0643
- Roy, J. S., MacDermid, J. C., & Woodhouse, L. J. (2010). A systematic review of the psychometric properties of the Constant-Murley score. *Journal of Shoulder and Elbow Surgery*, 19, 157-164. doi: 10.1016/j.jse.2009.04.008
- Saliba, I., Nader, M. E., El Fata, F., & Leroux, T. (2011). Bone anchored hearing aid in single sided deafness: Outcome in right-handed patients. *Auris Nasus Larynx*, 38, 570-576. doi: 10.1016/j.anl.2011.01.008
- Shiroma, M., Iwaki, T., Kubo, T., & Soli, S. (2008). The Japanese hearing in noise test. *International Journal of Audiology*, 47, 381-382. doi: 10.1080/14992020802054790
- Soli, S. D., & Wong, L. L. (2008). Assessment of speech intelligibility in noise with the hearing in noise test. *International Journal of Audiology*, 47, 356-361. doi: 10.1080/14992020801895136
- Streiner, D., & Norman, G. (2008). Reliability. In D. Streiner, & G. Norman (Eds.), *Health measurement scales: A practical guide to their development and use* (4th ed.). Oxford University Press, Oxford, UK. <http://www.oxfordscholarship.com/10.1093/acprof:oso/9780199231881.001.0001/acprof-9780199231881-chapter-8>
- Theunissen, M., Swanepoel, D. W., & Hanekom, J. (2009). Sentence recognition in noise: Variables in compilation and interpretation of tests. *International Journal of Audiology*, 48, 743-757. doi: 10.3109/14992020903082088
- Vaillancourt, V., Laroche, C., Giguère, C., Beaulieu, M. A., & Legault, J. P. (2011). Evaluation of auditory functions for Royal Canadian Mounted Police officers. *Journal of the American Academy of Audiology*, 22, 313-331. doi: 10.3766/jaaa.22.6.2
- Vaillancourt, V., Laroche, C., Giguère, C., & Soli, S. D. (2008). Establishment of age-specific normative data for the Canadian French version of the hearing in noise test for children. *Ear & Hearing*, 29, 453-466. doi: 10.1097/01.aud.0000310792.55221.0c
- Vaillancourt, V., Laroche, C., Mayer, C., Basque, C., Nali, M., Eriks-Brophy, A., ... Giguère, C. (2005). Adaptation of the HINT (hearing in noise test) for adult Canadian Francophone populations. *International Journal of Audiology*, 44, 358-369.
- Vermiglio, A. J. (2008). The American English hearing in noise test. *International Journal of Audiology*, 47, 386-387. doi: 10.1080/14992020801908251
- Vermiglio, A. J., Soli, S. D., Freed, D. J., & Fisher, L. M. (2012). The relationship between high-frequency pure-tone hearing loss, hearing in noise test (HINT) thresholds, and the articulation index. *Journal of the American Academy of Audiology*, 23, 779-788. doi: 10.3766/jaaa.23.10.4
- Wilson, R. H., McArdle, R. A., & Smith, S. L. (2007). An evaluation of the BKB-SIN, HINT, QuickSIN, and WIN materials on listeners with normal hearing and listeners with hearing loss. *Journal of Speech, Language and Hearing Research*, 50, 844-856. doi: 10.1044/1092-4388(2007)059
- Wolfe, J., Schafer, E. C., Heldner, B., Mulder, H., Ward, E., & Vincent, B. (2009). Evaluation of speech recognition in noise with cochlear implants and dynamic FM. *Journal of the American Academy of Audiology*, 20, 409-421.
- Wong, L. L., & Soli, S. D. (2005). Development of the Cantonese hearing in noise test (CHINT). *Ear & Hearing*, 26, 276-289.
- Wong, L. L., Soli, S. D., Liu, S., Han, N., & Huang, M. W. (2007). Development of the Mandarin hearing in noise test (MHINT). *Ear & Hearing*, 28, 70S-74S. doi: 10.1097/AUD.0b013e31803154d0

Remerciements

Les auteurs souhaitent remercier l'Institut de Réadaptation en déficience physique de Québec et le Centre Interdisciplinaire de recherche en réadaptation et intégration sociale pour leur soutien financier, cette étude ayant été réalisée grâce à une bourse d'études supérieures offerte à Mathieu Hotton par ces organismes. Les auteurs remercient également messieurs Luc Noreau et Jean-Sébastien Roy pour leurs commentaires sur les versions préliminaires de ce manuscrit.

Note des auteurs

Adresse pour correspondance : Mathieu Hotton,
Centre Interdisciplinaire de recherche en réadaptation
et intégration sociale, 525, boulevard Hamel, bureau
H-610, Québec (Québec), G1M 2S8 CANADA. Courriel :
mathieu.hotton.1@ulaval.ca.

ANNEXE 1

**Critical appraisal of study quality for psychometric articles
Interpretation Guide (Law & MacDermid, 2008)**

To decide which score to provide for each item on your quality checklist, read the following descriptors. Pick the descriptor that sounds most like the study you were evaluating with respect to a given item. If there is no documentation of an action, it is treated as not done.

Descriptors

Study question

Step/ Score

| | | |
|---|---|---|
| 1 | 2 | The authors: <ul style="list-style-type: none"> - Performed a thorough literature review indicating what is currently known about the psychometric properties of the instruments or tests under study from previous research studies - Presented a critical, and unbiased view of the current state of knowledge - Indicated how the current research question evolves from a gap in the current knowledge base - Established a research question based on the above. |
| | 1 | All of these above criteria were not fulfilled (little reference to previous research and present gaps in knowledge), but a clear rationale was provided for the research question. |
| | 0 | A foundation for the current research question was not clear and the rationale was not founded on previous literature. |

Study design

| | | |
|---|---|--|
| 2 | 2 | Specific inclusion/exclusion criteria for the study were defined, the practice setting was described and appropriate demographic information was presented yielding a study group generalizable to a clinical situation. |
| | 1 | Some information on participants and place is provided (NOT ALL). For example, age/sex/diagnosis and the name or type of the practice is given but not without additional information. Information on the type of patients is briefly defined, but it is insufficient to allow the reader to generalize the study to a specific population. |
| | 0 | No information on type of clinical settings or study participants is provided. |
| 3 | 2 | Authors identified specific hypotheses which included the specific type of reliability (intra/inter-rater or test-retest) or validity (construct/ criterion/ content; longitudinal/concurrent; convergent/ divergent) being tested. A prior hypothesis definitions of level of reliability and for validity, expected relationships (strength of associations) or constructs. |
| | 1 | Types of reliability and validity being tested were stated, but not clearly defined in terms of specific hypotheses. |
| | 0 | Specific types of reliability or validity under evaluation were not clearly defined nor were specific hypotheses on reliability and validity stated. (<i>"The purpose of this study was to investigate the reliability and validity of..."</i> can be rated it is zero if no further detail on the types of reliability and validity or the nature of specific hypotheses is stated). |

| | | |
|---|---|---|
| 4 | 2 | An appropriate scope of psychometric properties would be indicated by: <ol style="list-style-type: none"> 1. A detailed focus on reliability that included multiple forms of reliability (at least two of – intra-rater, inter-rater, test retest); as well as both relative and absolute reliability (e.g. ICCs and SEM/MID) 2. A detailed focus on validity that included multiple forms of validity (content- judgmental (structured e.g. expert review/survey or qualitative interviews) or statistical (e.g. factor analyses), construct (known group differences; convergent/divergent associations), criterion (concurrent/predictive), responsiveness; predictive, evaluative or discriminative properties were established 3. Some aspects of both reliability and validity were examined concurrently. |
| | 1 | Two or more psychometric properties were evaluated, however, scope was narrow/ and did not meet above criteria (e.g. internal consistency and one test/form of validity). |
| | 0 | The scope of psychometric properties was narrow as indicated by evaluation of only one form of reliability or validity. |
| 5 | 2 | Authors performed a sample size calculation and obtained their recruitment targets. Post-hoc power analyses and/or confidence intervals confirm that the sample size was sufficient to define relatively precise estimates of reliability or validity. |
| | 1 | The authors provide a rationale for the number of subjects included in the study, but did not present specific sample size calculations or post-hoc power analyses (or for simple reliability/validity statistics if sample >100 but no justification). |
| | 0 | Size of the sample was not rationalized or is clearly underpowered. |
| 6 | 2 | 90% or more of the patients enrolled for study were re-evaluated. |
| | 1 | More than 70% of the eligible patients were re-evaluated. |
| | 0 | Less than 70% of the patients eligible for study were re-evaluated. |

Measurements

| | | |
|---|---|---|
| 7 | 2 | Documentation of test procedures. The authors provided or referenced a published manual/article that outlines specific procedures for administration, scoring (including scoring algorithms handling of missing data) and interpretation that included any necessary information about positioning/active participation of the client, any special equipment required, calibration of equipment if necessary, training required, cost, examiner procedures/actions. If no manual then the text describes key details of procedures in sufficient detail they could be replicated. |
| | 1 | Procedures are referenced without any details or a limited description of procedures is included within text. |
| | 0 | Minimal description of procedures without appropriate references. |

| | | |
|---|---|---|
| 8 | 2 | Test performance in the study. All of the measurements, including test administration and scoring were performed in a standardized way. For self-report, this is characterized a statement of who administered the forms and by what process (definition of rules for exclusion of forms (e.g. too many missing items) or individuals (language, comprehension, etc.). For impairment measures, this would include calibration of any equipment; use of consistent measurement tools and scoring, a priori exclusion of any participants likely to give invalid results/unable to complete testing (not exclusion of after enrollment); use of standardized instructions and test procedures. |
| | 1 | No obvious sources of bias in how tests were performed/administered but minimal attention or description of the extent to which the above standards were maintained. |
| | 0 | No description of the extent to which the above standards were maintained or an obvious source of bias in data collection methods. |

Analyses

| | | |
|----|---|---|
| 9 | 2 | Authors clearly defined which specific analyses were conducted for each of the stated specific hypotheses of the study. This may be accomplished through organization of the results under specific subheadings or by demarcating which analyses addressed specific psychometric properties. Data was presented for each hypothesis/research question. |
| | 1 | Data was presented for each hypothesis, but authors did not clearly link analyses to hypotheses. |
| | 0 | Data was not presented for each hypothesis or psychometric property outlined in the purposes or methods. |
| 10 | 2 | Tests selected - Appropriate statistical tests were conducted: 1. Reliability (Relative=ICCs for quantitative, Kappa for nominal data); absolute (SEM or plot of score differences vs. average score showing mean and 2SD limit – Altman and Bland) 2. Clinical relevance - minimal detectable change, minimally important difference 3. Validity a. Validity associations - Pearson correlations for normally distributed data, Spearman rank correlations for ordinal data; or other correlations if appropriate b. Validity tests of significant difference- an appropriate global test like analysis of variance was used where indicated, with post-hoc tests that adjusted for multiple testing 3. Responsiveness - standardized response means or effect sizes or other recognized responsiveness indices were used. |
| | 1 | Data was presented for each hypothesis, but authors did not clearly link analyses to hypotheses. |
| | 0 | Data was not presented for each hypothesis or psychometric property outlined in the purposes or methods. |

| | | |
|----|---|---|
| 11 | 2 | <u>Range of analyses.</u> For key indicators like reliability coefficients indices, at least 2 of the following were presented: 1. appropriate confidence intervals, 2. Comparison to appropriate benchmarks or standards or 3. SEM. Correlation matrices for validity analysis may not require that each individual correlation be presented with its associated confidence intervals; however, confidence intervals and benchmarks should be used according to standards for that type of analysis. |
| | 1 | Either confidence intervals or appropriate benchmarks were used - not both. |
| | 0 | Inappropriate use of benchmarks or confidence intervals or neither included. |

Recommendations

| | | |
|----|---|--|
| 12 | 2 | Authors made specific conclusions and clinical recommendations that were clearly related to specific hypotheses stated at the beginning of the study and supported by the data presented. |
| | 1 | Authors made conclusions and clinical recommendations that were general but basically supported by the study data; OR authors made conclusions and clinical recommendations for only some of the study hypotheses. |
| | 0 | Authors made vague conclusions without any clinical recommendations; conclusions OR recommendations that were in contradiction to the actual data presented. |

ANNEXE 2

Sommaire de la méthodologie des études psychométriques incluses et des normes présentées

| Étude (version du HINT) | Population | n | Propriétés évaluées | Intervalle test-retest | Normes (M ± ET) |
|--|---|---|---------------------------------------|---|---|
| Baron de Otero, Brik, Flores, Ortiz & Abdala, 2008 (Espagnol latino-américain) | Locuteurs hispanophones latino-américains adultes ayant une acuité auditive normale, provenant de 14 pays latino-américains. | Préparation du matériel: 29 Estimation de la fonction PI: 18 (3 groupes de 6) Égalisation de la difficulté des phrases: 30 (3 groupes de 10) | Validité | N/A | N/A |
| Bevilacqua, Banhara, Da Costa, Vignoly & Alvarenga, 2008 (Portugais-brésilien) | Locuteurs du portugais brésilien, natifs du Brésil, âgés de 18 à 45 ans et ayant une acuité auditive normale. | Préparation du matériel: 10 Estimation de la fonction PI: 12 (2 groupes de 6) Égalisation de la difficulté des phrases: 20 (2 groupes de 10) Fidélité et normalisation: 29 | Validité Fidélité Normalisation | Test et retest réalisés lors de la même rencontre | <u>Sous écouteurs</u> <i>Silence (dBA)</i> 15.3 ± 3.5 <i>Bruit (dB S/B)</i> Face: -4.6 ± 0.8 Droite: -12.1 ± 1.0 Gauche: -12.2 ± 0.8 Composite: -8.4 ± 0.6 |
| Cekic & Sennaroglu, 2008 (Turc) | Locuteurs du turc, natifs de la Turquie, âgés de 17 à 45 ans et ayant une acuité auditive normale (seuils auditifs ≤ 15 dB HL de 250 à 8000Hz). | Préparation du matériel: 20 (2 groupes de 10) Estimation de la fonction PI: 7 Égalisation de la difficulté des phrases: 40 (4 groupes de 10) Création des listes de phrases: 12 Fidélité et normalisation: 30 | Validité Fidélité Normalisation | Test et retest réalisés lors de la même rencontre | <u>Sous écouteurs</u> <i>Silence (dBA)</i> 25.9 ± 2.2 <i>Bruit (dB S/B)</i> Face: -3.9 ± 0.9 Droite: -11.8 ± 1.1 Gauche: -12.1 ± 1.0 Composite: -7.9 ± 0.8 |

| | | | | | |
|---|--|---|--|---|---|
| Duncan & Aarts, 2006 (Anglais américain) | Locuteurs anglophones natifs des États-Unis, âgés de 22 à 29 ans et ayant une acuité auditive normale (seuils auditifs < 20 dB HL de 500 à 6000 Hz). | 15 | Fidélité | Test et retest réalisés lors de la même rencontre | N/A |
| Hallgren, Larsby & Arlinger, 2006 (Suédois) | Locuteurs du suédois, natifs de la Suède. Validité : - 19 à 40 ans - acuité auditive normale telle que rapportée par les sujets Fidélité et normalisation : - 18 à 30 ans - Acuité auditive normale (seuils ≤ 25 dB HL) | Préparation du matériel: 30 (2 groupes de 15) Égalisation de la difficulté des phrases: 96 (4 groupes de 24) Fidélité et normalisation: 10 | Validité Fidélité Normalisation | 1 semaine | <u>Sous haut-parleurs</u> <i>Bruit (dB S/B)</i> Face: -3.0 ± 1.1 |
| Hanks & Johnson, 1998 (Anglais américain) | Locuteurs anglophones natifs des États-Unis, âgés de 60 à 70 ans et ayant une surdité neurosensorielle légère (seuils auditifs < 40 dB HL en moyenne de 500 à 2000 Hz). | 24 | Fidélité | N/A | N/A |
| Huarte, 2008 (Espagnol castillan) | Locuteurs hispanophones natifs de l'Espagne, âgés de 20 à 50 ans et ayant une acuité auditive normale (seuils auditifs < 25 dB HL de 500 à 6000 Hz). | Préparation du matériel: 2 Estimation de la fonction PI: 10 Égalisation de la difficulté des phrases: 20 (2 groupes de 10) Création des listes de phrases: 30 Normalisation: 11 | Validité Normalisation préliminaire | N/A | <u>Sous écouteurs</u> <i>Silence (dBA)</i> 14.6 ± 3.2 <i>Bruit (dB S/B)</i> Face: -3.6 ± 1.2 Droite: -11.8 ± 1.2 Gauche: -12.1 ± 1.0 Composite: -7.7 ± 0.8 |

| | | | | | |
|---|--|---|---|---|---|
| Lamothe et al., 2002 (Anglais américain) | Sujets natifs du Canada : - 1 groupe d'anglophones unilingues - 1 groupe de francophones bilingues ayant appris l'anglais avant 11 ans - Âgés de 18 à 30 ans - Acuité auditive normale | 40, dont 20 anglophones unilingues et 20 francophones bilingues | Normalisation Applicabilité | Test et retest réalisés lors de la même rencontre | <u>Sous haut-parleurs</u> (étendue des normes fournies seulement) <i>Silence (dBA)</i> 8.57 à 20.64 <i>Bruit (dB S/B)</i> Face: -4.35 à 0.31 Droite: -13.21 à -6.36 Gauche: -11.40 à -4.53 |
| Laroche et al., 2006 (Français canadien pour enfants) | Locuteurs franco-canadiens natifs du Canada, ayant une acuité auditive normale (seuils auditifs \leq 15 dB HL de 500 à 4000Hz). Validité: - Enfants de 4.6 à 5.8 ans (m=5.0, SD=0.5) - Adultes âgés de 24 ans en moyenne (écart-type = 1,5 an). Fidélité et pré-normalisation: - Enfants de 6 à 9 ans - Adultes âge m=25 (3.1) | Validité : 10 enfants et 10 adultes Fidélité et pré-normalisation: 58 enfants et 9 adultes | Validité Fidélité Pré-normalisation | Test et retest réalisés lors de la même rencontre | N/A |
| Lolov, Raynov, Boteva & Edrev, 2008 (Bulgare) | Locuteurs du bulgare, natifs de Bulgarie, âgés de 15 à 38 ans et ayant une acuité auditive normale (seuils auditifs \leq 20 dB HL de 250 à 8000Hz). | Préparation du matériel: 5 Estimation de la fonction PI: 10 Égalisation de la difficulté des phrases: 20 (2 groupes de 10) Création des listes de phrases: 13 Fidélité et normalisation: 18 | Validité Fidélité Normalisation | Test et retest réalisés lors de la même rencontre | <u>Sous écouteurs</u> <i>Silence (dBA)</i> 24.2 \pm 7.5 <i>Bruit (dB S/B)</i> Face: -4.0 \pm 1.5 Droite: -9.7 \pm 3.2 Gauche: -10.2 \pm 3.9 Composite: -7.0 \pm 2.5 |

| | | | | | |
|---------------------------------|---|---|--|---|--|
| Moon et al., 2008 (Coréen) | Locuteurs du coréen, natifs de Corée, adultes et ayant une acuité auditive normale (seuils auditifs \leq 25 dB HL de 250 à 8000Hz). | Préparation du matériel: 5 Estimation de la fonction PI: 10 Égalisation de la difficulté des phrases: 40 (4 groupes de 10) Création des listes de phrases: 12 Fidélité et normalisation: 30 | Validité Fidélité Normalisation | Non disponible | <u>Sous écouteurs</u> <i>Silence (dBA)</i> 18.6 \pm 3.0 <i>Bruit (dB S/B)</i> Face: -3.3 \pm 1.0 Droite: -10.8 \pm 1.1 Gauche: -10.5 \pm 1.2 Composite: -7.0 \pm 0.7 |
| Myhrum & Moen, 2008 (Norvégien) | Locuteurs du norvégien, natifs de Norvège, adultes et ayant une acuité auditive normale (seuils auditifs \leq 25 dB HL de 125 à 8000Hz). | Préparation du matériel: 7 Estimation de la fonction PI: 9 Égalisation de la difficulté des phrases: 40 (2 groupes de 20) Création des listes de phrases: 7 Fidélité et normalisation: 60 | Validité Fidélité Normalisation | Test et retest réalisés lors de la même rencontre | <u>Sous écouteurs</u> <i>Silence (dBA)</i> 17.5 \pm 2.9 <i>Bruit (dB S/B)</i> Face: -3.2 \pm 1.0 Droite: -10.3 \pm 1.1 Gauche: -10.4 \pm 1.1 Composite: -6.7 \pm 0.8 |
| Nielsen & Dau, 2011 (Danois) | Locuteurs du danois, natifs du Danemark Groupe avec acuité auditive normale - Âgés entre 19 et 43 ans (m=33,6) - 50% femmes - seuils auditifs \leq 20 dB HL de 125 à 8000Hz - 1 seuil à 25 dB HL était toléré. Groupe avec surdit e - Âgés entre 61 et 69 ans (m=65,9) - 37,5% femmes - Surdit e NS l eg ere  a mod er ee (presbyacousie) - Exp erience 1 an avec port de p/a | Validit e: 12 Fid elit e et normalisation: 16 avec acuit e auditive normale et 16 avec surdit e | Validit e Fid elit e Normalisation | 3 semaines | <u>Sous  couteurs</u> <u>Sujets avec acuit e auditive normale</u> <i>Bruit (dB S/B)</i> Face: -2.52 \pm 0.87 <u>Sujets avec surdit e</u> <i>Bruit (dB S/B)</i> Face: 0.09 \pm 1.79 |

| | | | | | |
|---|---|--|---------------------------------------|---|---|
| Nilsson et al., 1994 (Anglais américain) | Locuteurs anglophones natifs des États-Unis, adultes et ayant une acuité auditive normale normale (seuils auditifs ≤ 15 dB HL de 250 à 8000Hz). Préparation du matériel et égalisation de la difficulté des phrases : - âge: 17-45 ans (m=24) Fidélité : - âge: 18 à 43 ans (m=26.8) | Préparation du matériel: 16 Égalisation de la difficulté des phrases: 78 Fidélité: 17 | Validité Fidélité | Test et retest réalisés lors de la même rencontre | N/A |
| Quar et al., 2008 (Malais) | Locuteurs du malais, natifs de Malaisie, adultes, avec acuité auditive normale Validité: - Âge: 20-40 ans Fidélité et normalisation : - Âge: 30-50 ans | Préparation du matériel: 6 Estimation de la fonction PI: 10 Égalisation de la difficulté des phrases: 20 (2 groupes de 10) Création des listes de phrases: 6 Fidélité et normalisation: 30 | Validité Fidélité Normalisation | Test et retest réalisés lors de la même rencontre | <u>Sous écouteurs</u> Silence (dBA) 20.2 \pm 3.4 Bruit (dB S/B) Face: -4.7 \pm 0.8 Droite: -12.4 \pm 1.0 Gauche: -12.4 \pm 1.0 Composite: -8.6 \pm 0.7 |
| Ribera, 2005 (Anglais américain, via télésanté) | Locuteurs anglophones natifs des États-Unis, âgés de 18 à 30 ans et ayant une acuité auditive normale. | Accord inter-juges: 20 Reproductibilité: 20 | Fidélité Applicabilité | Test et retest réalisés lors de la même rencontre | N/A |

| | | | | | |
|--|--|--|---------------------------------------|---|--|
| Shiroma et al., 2008 (Japonais) | Locuteurs du japonais, natifs du Japon, âgés entre 16 et 37 ans (m=22,5) et ayant une acuité auditive normale (seuils auditifs \leq 15 dB HL de 125 à 8000Hz). | Préparation du matériel: 12 Estimation de la fonction PI: 11 Égalisation de la difficulté des phrases: 9 Fidélité et normalisation : - Sous haut-parleurs: 85 - Écouteurs: 65 | Validité Fidélité Normalisation | Non disponible | <u>Sous écouteurs</u> <i>Silence (dBA)</i> 16.8 \pm 3.1 <i>Bruit (dB S/B)</i> Face: -5.3 \pm 1.4 Droite: -12.2 \pm 1.4 Gauche: -12.5 \pm 1.4 Composite: -8.8 \pm 1.1 |
| Vaillancourt et al., 2008 (Français canadien pour enfants) | Locuteurs franco-canadiens natifs du Canada, ayant une acuité auditive normale (seuils auditifs \leq 15 dB HL de 500 à 4000Hz). - Enfants de 6 à 12 ans - Adultes âgés de 18 à 30 ans | Développement de facteurs de correction en champ libre pour enfants normo-entendants : - 56 enfants (14 par groupe d'âge: 6, 8, 10 et 12 ans) - 14 adultes Fidélité : - 13 enfants - 15 adultes | Fidélité Normalisation | Test et retest réalisés lors de la même rencontre | Ne présentent pas de normes, mais des facteurs de correction applicables aux normes pour adultes développées pour un système de haut-parleurs donné. Permet la prédiction de normes en champ libre pour enfants à partir de normes pour adultes. |
| Vaillancourt et al., 2005 (Français canadien) | Locuteurs franco-canadiens natifs du Canada, âgés de 18 à 45 ans et ayant une acuité auditive normale (seuils auditifs \leq 15 dB HL de 250 à 8000Hz). Fidélité et normalisation : - 18 à 30 ans | Préparation du matériel: 9 Égalisation de la difficulté des phrases: 36 (groupes de 5, 7, 8, 6 et 10 sujets) Fidélité inter-listes: 9 Fidélité et normalisation: 36 | Validité Fidélité Normalisation | Test et retest réalisés lors de la même rencontre | <u>Sous écouteurs</u> <i>Silence (dBA)</i> 16.4 \pm 3.8 |

| | | | | | |
|--|--|--|------------------------|----------------|---|
| Vermiglio, 2008 (Anglais américain) | Locuteurs anglophones natifs des États-Unis, adultes âgés de 18 à 43 ans (m=38,7) et ayant une acuité auditive normale (seuils auditifs ≤ 15 dB HL de 250 à 8000Hz). | 67 | Fidélité Normalisation | Non disponible | <p><u>Sous-écouteurs</u></p> <p><i>Silence (dBA)</i> 15.6 ± 3.1</p> <p><i>Bruit (dB S/B)</i> Face: -2.6 ± 1.0 Droite: -10.1 ± 1.3 Gauche: -10.1 ± 1.3 Composite: -6.4 ± 0.9</p> |
| Vermiglio et al., 2012 (Anglais américain) | <p>Personnes ayant une acuité auditive normale (seuils auditifs de 2 à 6kHz ≤ 15 dB HL), âgées de 17 à 59 ans (m=33).</p> <p>Personnes avec surdité minimale (seuils auditifs de 2 à 6kHz entre 20-25 dB HL), légère (seuils auditifs de 2 à 6kHz entre 30-40 dB HL), modérée (seuils auditifs de 2 à 6kHz entre 45-55 dB HL), sévère (seuils auditifs de 2 à 6kHz entre 60-70 dB HL) et profonde (seuils auditifs de 2 à 6kHz entre 75-95 dB HL).</p> | <p>n total = 215</p> <p>Sujets avec acuité auditive normale - 51</p> <p>Sujets avec surdité</p> <ul style="list-style-type: none"> - Minimale: 56 - Légère: 63 - Modérée: 29 - Sévère: 12 - Profonde: 4 | Validité | N/A | N/A |

| | | | | | |
|---|---|---|---------------------------------------|---|---|
| Wilson et al., 2007 (Anglais américain) | <p>Locuteurs anglophones natifs des États-Unis</p> <p>Sujets avec acuité auditive normale</p> <p>- 18 à 30 ans (24.9 ± 2.8)</p> <p>- seuils auditifs ≤ 20 dB HL de 250 à 8000Hz</p> <p>Sujets avec surdit </p> <p>- 53 à 87 ans (70.8 ± 9.5)</p> <p>- MSP3: 34.5 dBHL ± 8.0</p> | <p>Sujets avec acuit  auditive normale</p> <p>- 24</p> <p>Sujets avec surdit </p> <p>- 72</p> <p>- MSP3: 2 sujets</p> <p>51-60 dB, 11 entre 41-50 dB, 37 entre 31-40 dB, 22 entre 20-30 dB.</p> | Validit  Fid lit  | Test et retest r alis s lors de la m me rencontre | N/A |
| Wong & Soli, 2005 (Cantonais) | <p>Locuteurs adultes du cantonnais natifs de Hong Kong, avec une acuit  auditive normale (seuils auditifs < 25 dB HL de 250 à 8000Hz).</p> | <p>Pr paration du mat riel: 9</p> <p>Estimation de la fonction PI: 3</p> <p> galisation de la difficult  des phrases:</p> <p>- Phase 1: 18 (3 groupes de 6)</p> <p>- Phase 2: 30 (5 groupes de 6)</p> <p>Fid lit  et normalisation:</p> <p>- 74 test s avec des listes de 10 phrases</p> <p>- 32 test s avec des listes de 20 phrases</p> | Validit  Fid lit  Normalisation | Test et retest r alis s lors de la m me rencontre | <p><u>Sous  couteurs</u></p> <p><i>Silence (dBA)</i> 19.4 ± 3.1</p> <p><i>Bruit (dB S/B)</i> Face: -4.0 ± 0.9 Droite: -10.9 ± 1.0 Gauche: -11.0 ± 0.8 Composite: -7.5 ± 0.7</p> |

| | | | | | |
|--|--|--|---------------------------------------|---|--|
| Wong et al., 2007 (Mandarin continental) | Locuteurs adultes du mandarin continental natifs de la Chine continentale, avec une acuité auditive normale (seuils auditifs < 25 dB HL de 250 à 8000Hz). Fidélité et normalisation : - Âge m=22.0 ans (SD=1.60) | Préparation du matériel: 16 Estimation de la fonction PI: 18 Égalisation de la difficulté des phrases: 24 (3 groupes de 8) Création des listes de phrases: 6 Fidélité et normalisation: 80 | Validité Fidélité Normalisation | Test et retest réalisés lors de la même rencontre | <u>Sous écouteurs</u> <i>Silence (dBA)</i> 14.7 ± 1.49 <i>Bruit (dB S/B)</i> Face: -4.3 ± 0.62 Droite: -11.7 ± 0.85 Gauche: -11.7 ± 0.79 Composite: -8.0 ± 0.57 |
| Wong et al., 2007 (Mandarin taiwanais) | Locuteurs adultes du mandarin taiwanais natifs de Taiwan, avec une acuité auditive normale (seuils auditifs < 25 dB HL de 250 à 8000Hz). | Préparation du matériel: 16 Estimation de la fonction PI: 18 Égalisation de la difficulté des phrases: 24 (3 groupes de 8) Création des listes de phrases: 6 Fidélité et normalisation: 32 | Validité Fidélité Normalisation | Test et retest réalisés lors de la même rencontre | <u>Sous écouteurs</u> <i>Silence (dBA)</i> 19.4 ± 3.1 <i>Bruit (dB S/B)</i> Face: -4.0 ± 0.94 Droite: -10.9 ± 1.0 Gauche: -11.0 ± 0.8 Composite: -7.5 ± 0.72 |