CANADIAN JOURNAL OF SPEECH-LANGUAGE PATHOLOGY & AUDIOLOGY | CJSLPA

Volume 43, No. 1, 2019

REVUE CANADIENNE D'ORTHOPHONIE ET D'AUDIOLOGIE | RCOA

Volume 43, No. 1, 2019

SAC OAC

Speech-Language & Audiology Canada

Orthophonie et Audiologie Canada

Communicating care
La communication à coeur

CJSLPA REVIEWERS

Reviewers for this issue included: Marie-Ève Caty, Monique Charest, Christelle Maillart, Kathleen Peets, Stefano Rezzonico, and Catherine Wiseman-Hakes.

VISION AND MISSION OF SPEECH-LANGUAGE AND AUDIOLOGY CANADA

VISION
Ensuring all people of Canada achieve optimal communication health.

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

INDEXING

CJSLPA is indexed by:

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

ISSN 1913-2019

SCOPE AND PURPOSE OF CJSLPA

SCOPE

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

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

PURPOSE

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

COPYRIGHT

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

RÉVISEURS DE LA RCOA

Les personnes suivantes ont agi à titre de réviseurs pour ce numéro : Marie-Ève Caty, Monique Charest, Christelle Maillart, Kathleen Peets, Stefano Rezzonico et Catherine Wiseman-Hakes.

VISION ET MISSION D'ORTHOPHONIE ET AUDIOLOGIE CANADA

VISION

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

MISSION

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

INDEXATION

La RCOA est indexée dans :

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

ISSN 1913-2019

MISSION ET BUT DE LA RCOA

MISSION

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

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

BUT

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

DROIT D'AUTEUR

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

TABLE OF CONTENTS

TABLE DES MATIÈRES

La compréhension syntaxique des enfants de maternelle ayant un trouble développemental du langage : phrases simples et complexes

Syntactic Comprehension of Kindergarten French-Speaking Children with Developmental Language Disorder: Simple and Complex Sentences

MOTS-CLÉS
TROUBLE DÉVELOPPEMENTAL DU LANGAGE
COMPRÉHENSION
SYNTAXE

Élody Ross-Lévesque
Chantal Desmarais
Caroline Arsenault
Geneviève Demers-Jacques

Abrégé

Les enfants ayant un trouble développemental du langage ont un déficit marqué à maîtriser la grammaire de leur langue, ce qui les place à risque d'avoir des difficultés communicatives dans plusieurs sphères de leur vie. Or, peu de données franco-québécoises sont actuellement disponibles sur la compréhension syntaxique des enfants ayant un trouble développemental du langage. Cette étude vise à explorer les habiletés de compréhension de phrases simples et complexes des enfants franco-québécois de maternelle ayant un trouble développemental du langage. Une tâche de compréhension de phrases a été utilisée. Les résultats montrent que les enfants de maternelle ayant un trouble développemental du langage ont plus de difficultés à comprendre les phrases complexes que les enfants au développement typique, notamment les phrases relatives, passives, avec un pronom en position objet et les questions avec le pronom interrogatif « qui ».

Élody Ross-Lévesque et Chantal Desmarais

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

Caroline Arsenault et Geneviève Demers-Jacques

Université Laval, Québec, QC, CANADA

Abstract

Children with developmental language disorder have difficulties with the grammatical features of language. This creates communicative challenges in many aspects of their lives. There is limited data on the syntactic comprehension of French-speaking children with developmental language disorder. This study explores the comprehension of simple and complex sentences among francophone kindergarten children with developmental language disorder. A sentence comprehension task was used. Results showed that kindergarten children with developmental language disorder have greater difficulty understanding syntactically complex sentences compared to typically developing children, especially relative, passive, pronominal sentences and wh-questions.

Les enfants ayant un trouble développemental du langage (TDL), soit environ 7% des enfants qui fréquentent la maternelle (Tomblin et al., 1997), sont reconnus pour avoir des difficultés significatives et persistantes dans l'acquisition et la maîtrise du langage, tant sur le plan expressif que réceptif, et ce, en l'absence de déficit neurologique, auditif ou intellectuel (Bishop, Snowling, Thompson, Greenhalgh et CATALISE consortium, 2016; Leonard, 2014b). Plus spécifiquement, de nombreuses études ont montré que les enfants ayant un TDL éprouvent de la difficulté à maîtriser les aspects morphologiques et syntaxiques de leur langue (Leonard, 2014b; Montgomery, 2003; Montgomery, Gillam et Evans, 2016; Reilly et al., 2014). Les difficultés langagières de ces enfants, malgré les améliorations dues à l'intervention et à la maturation, les placent à fort risque de vivre des difficultés à long terme, et ce, tant au plan social qu'émotionnel et académique (Botting, Durkin, Toseeb, Pickles et Conti-Ramsden, 2016; Johnson, Beitchman et Brownlie, 2010; Leonard, 2014a). Il s'agit donc d'une problématique développementale identifiée pendant la petite enfance qui a des répercussions tout au long de la vie.

Or, peu de données sont disponibles quant à l'acquisition de la syntaxe des enfants francophones ayant un TDL. De plus, les données sur l'acquisition de la syntaxe des enfants ayant un TDL décrivent davantage les habiletés expressives que réceptives (Montgomery et al., 2016). Pourtant, une meilleure connaissance des habiletés syntaxiques réceptives est essentielle pour une prise en charge efficace des enfants ayant des difficultés de compréhension. D'abord, il est reconnu que les enfants ayant des déficits dans les sphères expressives et réceptives sont plus à risque de difficultés au plan académique que les enfants présentant des déficits expressifs seulement (Conti-Ramsden, Durkin, Simkin et Knox, 2009; Justice, Bowles, Pence Turnbull et Skibbe, 2009). De plus, étant donné l'absence de bases théoriques solides, peu d'interventions ayant pour objectif de stimuler la compréhension syntaxique des enfants ayant un TDL ont été étudiées (Ebbels, 2007). La présente étude s'inscrit donc dans un créneau peu exploré à ce jour, soit la compréhension de phrases d'enfants francophones qui présentent un TDL.

Étant donné l'importance des habiletés grammaticales dans le profil communicatif des enfants ayant un TDL, il importe de pouvoir en dégager un portrait exhaustif. Pour mettre en évidence la présence de difficultés sur le plan de l'acquisition de la grammaire, il est incontournable d'avoir au préalable une représentation claire de ce qui constitue un développement typique (DT). Il est également important de savoir que les enfants produisent fréquemment certains types de phrases qu'ils ne comprennent pas encore parfaitement (Paul et Norbury, 2012). Ainsi, les jalons développementaux expressifs ne suffisent pas pour fournir un portrait exhaustif des habiletés syntaxiques. À la lumière des données de recherche actuellement disponibles, il est possible de dresser un portait des habiletés réceptives syntaxiques des enfants francophones et non francophones au DT et ayant un TDL. Il faut toutefois noter que peu de données sont disponibles pour l'acquisition du français, une lacune que cette étude tente de combler.

Développement de la compréhension syntaxique des enfants francophones au DT et ayant un TDL

Leonard (2014b) s'est intéressé aux manifestations du TDL chez des enfants qui parlent différentes langues. Une des constatations clés qui se dégage des études répertoriées est que les erreurs syntaxiques des enfants ayant un TDL sont également observées chez les enfants au DT et correspondent aux éléments difficiles de leur langue d'usage. Les enfants ayant TDL différeraient donc de leurs pairs par leur lenteur d'acquisition et par leur vulnérabilité aux éléments difficiles de leur langue (Leonard, 2014b). Lorsqu'on s'intéresse spécifiquement au développement syntaxique des enfants francophones, on note que très peu d'études sont disponibles et celles-ci ont été respectivement réalisées auprès d'enfants francophones de la France, de la Belgique et de la Suisse. Chez les enfants francophones au DT, les phrases actives seraient typiquement comprises dès l'âge de 43 mois, alors que les phrases passives non-réversibles et les relatives avec « qui » seraient acquises entre 43 mois et 56 mois (Segui et Léveillé, 1977). Deux autres études plus récentes se sont penchées sur la compréhension de phrases complexes. L'une d'elles a notamment observé la compréhension syntaxique de phrases complexes chez des enfants au DT âgés entre 5 et 12 ans. Pour ce faire, les auteurs ont utilisé une tâche standardisée de désignation d'images qui comprenaient des phrases impliquant un mouvement syntaxique (p. ex. Le mouton est poursuivi par le garçon) ou l'enchâssement d'une proposition dans une autre (p. ex. La vache que le chien poursuit est marron). Cette étude a montré que les habiletés de compréhension syntaxique continuent d'augmenter avec l'âge et qu'il y a des différences significatives entre les performances des enfants âgés de 5-6 ans, de 8-9 ans et de 11-12 ans (Delage et Frauenfelder, 2012). La compréhension de phrases et l'effet de la longueur des énoncés ont également été mesurés chez des enfants francophones âgés entre 8 et 13 ans, avec ou sans TDL (Leclercq, Majerus, Prigent et Maillart, 2013). Pour ce faire, les auteurs ont utilisé une tâche de

désignation d'images incluant des phrases courtes[1] (7 à 9 syllabes) et des phrases longues[2] (15 à 17 syllabes). Les résultats ont montré que la longueur des phrases a un effet sur les performances et sur la vitesse de réponse, et ce, tant chez les enfants ayant un TDL que chez les enfants au DT appariés sur l'âge ou sur leurs habiletés langagières.

Ceci dit, aucune étude n'a mesuré la compréhension de phrases complexes chez des enfants francophones plus jeunes (c.-à-d. âgés de cinq ans et qui fréquentent la maternelle). Ainsi, bien que le développement du langage en français diffère de celui dans les autres langues, il est utile de dresser un portrait des habiletés syntaxiques des enfants non francophones, d'autant plus que cette composante a été davantage étudiée chez les enfants anglophones. La description du développement des habiletés réceptives des enfants non francophones au DT ou ayant un TDL combinée aux données disponibles chez les enfants francophones permettra d'avoir une vision globale des connaissances actuelles.

Développement de la compréhension syntaxique des enfants non francophones au développement normal et ayant un TDL

La majorité des études qui portent sur la compréhension syntaxique ont utilisé une tâche de désignation d'images auprès d'enfants anglophones. Entre l'âge de 2 et 3 ans, les enfants au DT comprennent les phrases actives réversibles probables[3]. Toutefois, ils sont encore susceptibles d'interpréter les phrases réversibles improbables[4] en fonction de ce qui est le plus probable, et ainsi, d'attribuer à tort les rôles d'agent et de patient. Entre l'âge de 3 et 4 ans, la compréhension s'est normalement assez développée pour rendre l'enfant moins susceptible d'effectuer une interprétation erronée des phrases actives improbables, ce qui lui permettrait une meilleure compréhension des phrases actives probables et improbables (Evans, 2002). Entre l'âge de 4 et 5 ans, les enfants ne seraient plus du tout influencés par la probabilité de la phrase et auraient ainsi une pleine compréhension des phrases actives réversibles improbables (Paul et Norbury, 2012). Chez les enfants ayant un TDL, la difficulté à maîtriser les éléments grammaticaux de la langue est l'un des éléments saillants du trouble (Bishop, Bright, James, Bishop et van der Lely, 2000; Leonard, 2014b). Il a donc été proposé que les stratégies de probabilité utilisées chez les enfants plus jeunes et pouvant entraîner une compréhension erronée seraient susceptibles d'être utilisées plus longtemps par les enfants ayant un TDL, en particulier lorsque la demande cognitive augmente (Evans, 2002). Ainsi, il est attendu que le développement des habiletés de compréhension syntaxique chez les enfants ayant un TDL diffère de celui des enfants au DT.

Des difficultés de compréhension ont été identifiées chez les enfants ayant un TDL à l'aide de tâches de désignation d'images pour quatre types de structures syntaxiques : les phrases relatives, les phrases passives, les phrases avec un pronom en position objet et les questions avec le pronom interrogatif « qui » (Montgomery et al., 2016). Selon les différentes études répertoriées, ces types de phrases s'avèrent plus difficiles à comprendre en raison du déplacement de syntagme. Les enfants ayant un TDL obtiendraient donc des résultats plus faibles que les enfants au DT du même âge et que les enfants plus jeunes appariés en fonction des habiletés langagières.

Chez les enfants au DT, les phrases relatives seraient comprises entre l'âge de 4 et 7 ans, tout dépendant du degré de complexité de la phrase (de Villiers, Tager Flusberg, Hakuta et Cohen, 1979). Chez les enfants ayant un TDL, la difficulté à comprendre les phrases relatives serait liée au caractère non canonique de ce type de phrases. Ainsi, les enfants ne peuvent pas se fier à l'ordre des mots pour en comprendre le sens et doivent traiter le mouvement syntaxique (Montgomery et al., 2016). L'étude de Montgomery (2004) a étudié la compréhension d'enfants ayant un TDL âgés entre 6 et 10 ans pour différents types de phrases : phrases actives avec indice numérique[5], phrases relatives sujets[6], phrases avec doubles relatives sujets et objets[7] et phrases actives avec traitement des adjectifs[8]. Les résultats ont montré que les phrases relatives seraient plus difficiles à comprendre pour les enfants ayant un TDL que pour les enfants au DT et les enfants plus jeunes appariés en fonction des habiletés de compréhension syntaxique.

Tout comme les phrases relatives, les phrases passives sont de nature non canonique. Ainsi, un traitement du mouvement syntaxique est nécessaire pour bien comprendre la phrase. Deux études ont comparé la

[1] P. ex. La madame voit le garçon qui glisse.

[2] P. ex. Ce soir, la belle dame noire appelle la petite fille qui lit dans le pré.

[3] P. ex. La maman nourrit le bébé.

[4] P. ex. Le bébé nourrit la mère.

[5] P. ex. *Point to the picture of the three cats* (« Pointe l'image des trois chats »).

[6] P. ex. *The girl who is smiling is pushing the boy* (« La fille qui sourit pousse le garçon »).

[7] P. ex. *The little boy who is standing is hitting the little girl who is sitting* (« Le petit garçon qui est debout frappe la petite fille assise »).

[8] P. ex. *The dirty little boy climbed the big tall tree* (« Le petit garçon sale a escaladé le gros et grand arbre »).

compréhension de phrases passives[9] et de phrases actives[10] chez des enfants au DT et ayant un TDL âgés entre 6 et 13 ans (Bishop et al., 2000; Montgomery et Evans, 2009). Comme pour la compréhension de phrases relatives, les résultats indiquent que les enfants ayant un TDL présentent des difficultés de compréhension des phrases passives qui sont normalement acquises entre l'âge de 5 et 6 ans chez les enfants au DT (Bishop et al., 2000; Montgomery et Evans, 2009; Montgomery, Magimairaj et O'Malley, 2008; Paul et Norbury, 2012).

Ces mêmes études ont mesuré la compréhension de phrases avec un pronom en position objet[11] chez des enfants ayant un TDL (Bishop et al., 2000; Evans, 2002). Leurs résultats ont confirmé que les phrases avec un pronom en position objet, habituellement comprises entre l'âge de 5 et 6 ans chez les enfants au DT, sont plus difficilement comprises par les enfants ayant un TDL que par les enfants au DT appariés sur l'âge (Bishop et al., 2000; Montgomery et Evans, 2009; Montgomery et al., 2008; Paul et Norbury, 2012). Des difficultés seraient également constatées pour les questions avec le pronom interrogatif « qui ». Selon l'hypothèse *The Deficit in Computational Grammatical Complexity* (van der Lely, 2004, 2005; van der Lely, Rosen et McClelland, 1998), les enfants ayant un TDL traiteraient les déplacements syntaxiques de façon optionnelle plutôt qu'obligatoire. Ainsi, ces enfants auraient des difficultés à comprendre des phrases impliquant des mouvements syntaxiques entre la forme sous-jacente et la forme de surface (p. ex. les questions objets; Lin, 2006; Marinis et van der Lely, 2007). L'étude de Deevy et Leonard (2004) a testé cette hypothèse en mesurant la compréhension des questions avec le pronom interrogatif « qui » à l'aide de quatre catégories de questions : questions sujets courtes[12], questions sujets longues[13], questions objets courtes[14] et questions objets longues[15]. Les résultats ont montré que les enfants ayant un TDL âgés entre 4 et 6 ans présentaient plus de difficultés avec les questions longues et syntaxiquement complexes (p. ex. les questions objets) que les enfants au DT appariés en fonction du vocabulaire réceptif. Les questions courtes (sujets et objets) et les questions sujets seraient bien comprises par les enfants au DT et ayant un TDL.

Outre les difficultés en lien avec la structure syntaxique, la longueur des phrases pourrait avoir un effet important sur la compréhension des enfants ayant un TDL. En ce sens, Montgomery (1995, 2000a, 2000b, 2004) a montré à plusieurs reprises que les enfants ayant un TDL âgés entre 5 et 10 ans ont plus de difficultés de compréhension des phrases longues que les enfants au DT du même âge et que les enfants plus jeunes appariés en fonction des habiletés langagières réceptives et/ou expressives. Les phrases courtes seraient, quant à elles, bien comprises par tous les enfants. L'étude de Montgomery (2004), a analysé la compréhension de phrases courtes et longues. Selon cette étude, les phrases longues seraient significativement plus difficiles à comprendre que les phrases courtes pour les enfants ayant un TDL âgés entre 6 et 10 ans, lorsque comparés à des enfants au DT appariés sur l'âge et sur les habiletés de compréhension syntaxique. L'effet de longueur a également été constaté chez des enfants de différentes langues, dont le portugais et l'italien (Fortunato-Tavares et al., 2012; Pettenati, Benassi, Deevy, Leonard et Caselli, 2015). Une étude de Leonard, Deevy, Fey et Bredin-Oja (2013) a toutefois montré, à l'aide d'une tâche de désignation d'images, que ce serait plutôt la complexité sur le plan cognitif qui aurait un effet sur la performance et non la longueur de la phrase. En effet, dans cette étude, les enfants ayant un TDL âgés entre 4 et 6 ans présentaient des difficultés de compréhension avec les phrases actives réversibles courtes et longues[16], lorsque comparés à des enfants au DT du même âge. Toutefois, les difficultés étaient particulièrement marquées pour les phrases cognitivement plus complexes qui demandent un traitement spécifique des adjectifs[17]. Les enfants ayant un TDL présentaient en fait des résultats semblables aux enfants au DT plus jeunes appariés en fonction des habiletés de compréhension syntaxique. Ainsi, selon ces auteurs, ce serait les ressources cognitives des enfants ayant un TDL, et non seulement la structure des phrases, qui influenceraient les capacités de compréhension.

Les données disponibles au sujet de la compréhension de phrases des enfants anglophones, tant ceux au DT que ceux ayant un TDL, permettent d'émettre des hypothèses soutenant l'intervention en orthophonie. Par contre, il

[9]P. ex. *The woman is kissed by the baby* (« La femme est embrassé par le bébé »).
[10]P. ex. *The clown is hugging the tiny white elephant* (« Le clown étreint le petit éléphant blanc »).
[11]P. ex. *Bugs Bunny says Daffy Duck is hugging him* (« Bugs Bunny dit que Daffy Duck l'étreint »).
[12]P. ex. *Who is feeding the tiger?* (« Qui est en train de nourrir le tigre? »).
[13]P. ex. *Who is feeding the big orange tiger?* (« Qui est en train de nourrir le gros tigre orange »).
[14]P. ex. *Who is the bunny touching?* (« Qui le lapin est-il en train de touché? »).
[15]P. ex. *Who is the happy white bunny touching?* (« Qui le joyeux lapin blanc est-il en train de touché? »).
[16]P. ex. *The nice mouse covers the pretty bird* (« La gentille souris protège le joli oiseau »).
[17]P. ex. *The yellow dog washes the white pig* (« Le chien jaune lave le cochon blanc »).

y a un manque de connaissances quant aux habiletés syntaxiques réceptives des enfants francophones de maternelle, tant ceux au DT que ceux ayant un TDL. En fait, une revue approfondie de la littérature n'a permis d'identifier aucune étude qui rapporte des observations quant à la compréhension syntaxique de ces enfants. Il peut donc s'avérer ardu pour les intervenants travaillant auprès de cette population de connaître les éléments typiquement difficiles pour ces enfants, et ainsi, d'être en mesure de répondre aux besoins de ceux-ci en intervention. Ainsi, une meilleure connaissance des habiletés réceptives des enfants francophones ayant un TDL permettrait aux intervenants de mieux aider ces enfants dans le développement de leur langage, et ainsi, prévenir certaines difficultés à long terme.

Objectifs de l'étude

La présente étude explore les habiletés de compréhension de phrases simples et complexes d'enfants francophones.

1. Le premier objectif est de comparer la compréhension syntaxique d'enfants québécois de maternelle ayant un TDL à celle d'enfants au DT. Il est attendu que les enfants ayant un TDL obtiennent de plus faibles performances en compréhension de phrases complexes (c.-à-d. phrases relatives, phrases passives, phrases avec un pronom en position objet, questions objets et questions longues) que les enfants au DT du même âge, alors qu'il est attendu que les phrases simples (phrases actives, phrases avec un nom en position objet et questions sujets courtes) soient bien comprises par tous les enfants (Evans, 2002; Montgomery et al., 2016). En effet, il a été montré que les phrases impliquant un déplacement de syntagme et les phrases longues demandent davantage de temps pour être comprises par les enfants au DT et représentent donc un défi pour les enfants ayant un TDL. Les performances des enfants ayant un TDL devraient être similaires, voire plus faibles, que celle des enfants au DT plus jeunes appariés en fonction des habiletés langagières (Deevy et Leonard, 2004; Leonard et al., 2013; Montgomery, 2004).

2. Le deuxième objectif est d'analyser les effets de la complexité syntaxique et de la longueur sur la performance en compréhension de phrases des enfants. Selon Montgomery et ses collaborateurs (2016), la compréhension syntaxique des enfants ayant un TDL devrait être affectée par la voix passive, la catégorie grammaticale en position objet (nom ou pronom), par le déplacement de syntagme (questions objets) et la longueur des questions (questions longues).

3. Le troisième objectif est de déterminer les structures syntaxiques acquises et non acquises chez les enfants québécois au DT âgés entre 4 et 5 ans. Il est attendu que les phrases simples, normalement acquises entre l'âge de 3 et 4 ans, soient comprises par tous les enfants (Evans, 2002). Toutefois, les phrases relatives, les phrases passives et les phrases avec un pronom en position objet pourraient ne pas être pleinement maîtrisées par les enfants au DT âgés entre 4 et 5 ans (de Villiers et al., 1979; Montgomery et al., 2008; Paul et Norbury, 2012). Étant donné le peu de littérature portant sur le développement syntaxique des enfants québécois, les hypothèses s'appuient sur les études portant sur le développement syntaxique des enfants anglophones.

Méthodologie

L'étude présente un sous-ensemble des données recueillies dans le cadre d'une plus vaste étude qui a été approuvée par le comité d'éthique et de recherche de l'Institut de réadaptation en déficience physique de Québec (IRDPQ; #2008-132).

Participants

Soixante-sept enfants (35 garçons et 32 filles) provenant de la ville de Québec et de ses environs ont été recrutés pour cette étude. Trois groupes ont été formés afin de répondre aux objectifs de l'étude. Le groupe expérimental comprend 16 enfants de maternelle ayant un TDL (âge moyen = 5;3 ans). Ces enfants devaient répondre aux critères qui déterminent l'accès aux services du programme Déficience de langage de l'IRDPQ (c.-à-d. avoir une conclusion orthophonique de TDL, ou encore, d'hypothèse de TDL), être suivi en orthophonie au moment de l'étude et avoir le français comme langue maternelle. Dans ce contexte, le TDL se définit par une atteinte au moins modérée dans au moins une sphère du langage réceptif ou expressif. Deux groupes contrôles d'enfants au DT ont aussi été inclus dans l'étude : un groupe de 26 enfants (âge moyen = 5;1 ans) au DT appariés en fonction du niveau scolaire (DT-maternelle) et un groupe de 25 enfants (âge moyen = 4;7 ans) au DT appariés en fonction du niveau de vocabulaire réceptif (DT-préscolaire), soit leur score brut à l'Échelle de vocabulaire en images Peabody (ÉVIP; Dunn, Dunn et Thériault-Whalen, 1993). Ces deux groupes permettront de mieux comprendre la trajectoire développementale des enfants ayant un TDL et de mieux observer l'ampleur de leurs difficultés. Afin de vérifier que le groupe DT-préscolaire était bien apparié au groupe d'enfants ayant un TDL, une ANOVA a été réalisée et a confirmé qu'il n'existait aucune différence significative entre ces groupes quant à leur performance à l'ÉVIP ($p > 0,05$). Des tests de Fisher ont été réalisés sur les différentes

Tableau 1. Caractéristiques des participants

	TDL (*n* = 16)	DT-maternelle (*n* = 26)	DT-préscolaire (*n* = 25)
Sexe (M;F)	8;8	12;14	15;10
Âge moyen (ÉT)	5,3 (0,4)	5,0 (0,2)	4,7 (0,6)
ÉVIP-moyenne des scores bruts (ÉT)	55,50 (17,9)	78,50 (15,6)*	50,04 (17,5)
Scolarité du parent répondant			
Aucun risque (%)	50,0*	96,2	96[3]
À risque[1] (%)	50,0*	3,8	0
Revenu familial			
Aucun risque (%)	62,5*	96,2	88,0
À risque[2] (%)	37,5*	3,8	12,0

Note. [1]Dans cette étude, avoir un niveau de scolarité inférieur à un secondaire 5 est considéré comme étant à risque. [2]Dans cette étude, avoir un revenu familial inférieur à 40 000$ est considéré comme étant à risque. [3]Un des répondants n'a pas indiqué son niveau de scolarité. *Différences significatives avec les deux autres groupes (p < 0,05). ÉVIP = Échelle de vocabulaire en images Peabody; DT-maternelle = groupe d'enfants au développement typique appariés sur l'âge; DT-préscolaire = groupe d'enfants au développement typique appariés sur les habiletés de vocabulaire réceptif; ÉT = écart-type; F = féminin; M = masculin; TDL = groupe d'enfants ayant un trouble développemental du langage.

variables dichotomiques et montrent que les trois groupes à l'étude sont équivalents en ce qui concerne leur proportion de garçons et de filles (p = 0,640). Des différences sont observées quant à la scolarité du répondant (p < 0,001) et du revenu familial (p = 0,019). Le tableau 1 présente différentes caractéristiques des participants.

Le groupe d'enfants ayant un TDL a été recruté avec l'aide des orthophonistes de l'IRDPQ et via une étude antérieure. Les enfants des groupes contrôles ont été recrutés via la même étude antérieure, ainsi que par l'intermédiaire d'une liste de courriels des étudiants et employés de l'Université Laval. Un questionnaire parental a confirmé que les enfants des groupes contrôles ne présentaient pas de difficultés langagières ou développementales et avaient le français comme langue maternelle.

Matériel

Les habiletés de compréhension ont été évaluées à l'aide d'une tâche de compréhension orale de 56 phrases (pour détails, voir l'annexe) tirées de l'*Épreuve de Compréhension Syntaxico-Sémantique* (n = 32) et d'une tâche expérimentale de compréhension de questions (n = 24). Les 56 phrases ont été réparties aléatoirement. La tâche comprend 20 phrases simples, soit des phrases actives (p. ex. L'homme poursuit le chien), des phrases avec un nom en position objet (p. ex. Le monsieur arrose les fleurs) et des questions sujets de structure canonique (p. ex. Qui est en train de regarder le cheval?), et 36 phrases complexes sous-divisées en fonction des catégories mises de l'avant par Montgomery et ses collaborateurs (2016) : phrases relatives (p. ex. Le garçon mange les pommes que la fille cueille), phrases passives (p. ex. Le mouton est poursuivi par le garçon), phrases avec un pronom en position objet (p. ex. La vache les regarde) et questions complexes avec le pronom interrogatif « qui » (p. ex. Qui est en train d'arroser le petit cochon blanc?). Dans le cadre de cette étude, seules des phrases relatives avec le pronom « que » ont été utilisées. La tâche expérimentale de compréhension de questions est une adaptation française de la tâche utilisée dans l'étude

de Deevy et Leonard (2004). Elle comprend quatre types de questions avec le pronom interrogatif « qui » : les questions sujets courtes (p . ex. Qui est en train de regarder le cheval?), les questions sujets longues (p. ex. Qui est en train d'arroser le petit cochon blanc?), les questions objets courtes (p. ex. Qui l'oiseau est-il en train de poursuivre?) et les questions objets longues (p. ex. Qui le petit cochon blanc est-il en train d'arroser?). Les questions longues ont la même structure que les courtes, mais comprennent en plus deux adjectifs comptant d'une à deux syllabes. La périphrase « en train de » a été utilisée dans la formulation des questions afin d'éviter la présence d'un pronom sujet en fin de question (p. ex. Qui le cheval regarde-t-il?). Tel que mentionné précédemment, les questions sujets courtes sont incluses dans la catégorie des phrases simples étant donné qu'elles ne comprennent pas de déplacement de syntagme. Les questions complexes, soient les questions objets (courtes et longues) et les questions sujets longues sont quant à elles incluses dans la catégorie des phrases complexes. Le tableau 2 décrit les différentes caractéristiques des phrases utilisées.

La tâche de compréhension orale de phrases comprend 56 planches de quatre images chacune. Chaque planche comportait l'image cible et trois distracteurs : (1) image représentant la même action et les mêmes agents/patients que la cible, mais avec les rôles d'agent et de patient inversés, (2) image représentant une action différente de la cible et (3) image représentant la même action que la cible, mais réalisée par un agent/patient différent par le type d'animal ou le sexe de la personne. Par exemple, les distracteurs de l'item « La fille pousse le cheval » sont (1) « Le cheval pousse la fille », (2) « La fille est assise sur le cheval » et (3) « La fille pousse le garçon ». Pour les questions longues, les deux premiers distracteurs étaient identiques. Le troisième distracteur a été modifié pour inclure le traitement des adjectifs. Celui-ci illustrait donc les mêmes agents/patients et la même action que la cible. Cependant, étant donné qu'une même planche servait à présenter une question longue objet et une question longue sujet et que chaque planche ne contenait que quatre images, soit les rôles d'agent et de patient étaient inversés, soit une caractéristique (p. ex. la couleur) différenciait un des animaux par rapport à la cible. Par exemple, les distracteurs de la question longue « Qui est en train d'arroser le petit cochon blanc ? » sont (1) « Qui le petit cochon blanc est-il en train d'arroser? », (2) « Qui est en train d'embrasser le petit cochon blanc? » et (3) « Qui est en train d'arroser le petit cochon rose? ».

Tableau 2. Caractéristiques des stimuli

Types de phrases	Exemples	*n*
Simples		
Actives	L'homme poursuit le chien.	6
Nom en position objet	La fille laisse tomber la tasse.	8
Questions sujets courtes	Qui est en train de regarder le cheval?	6
Complexes		
Relatives	Le garçon mange les pommes que la fille cueille.	4
Passives	Le mouton est poursuivi par le garçon.	6
Pronom en position objet	La vache les regarde.	8
Questions sujets longues	Qui est en train d'arroser le petit cochon blanc?	6
Questions objets courtes	Qui l'oiseau est-il en train de poursuivre?	6
Questions objets longues	Qui le petit cochon blanc est-il en train d'arroser?	6

Procédures

La batterie d'évaluation, incluant la tâche de compréhension, a été administrée au Centre interdisciplinaire de recherche en réadaptation et intégration sociale en présence de l'enfant, du/des parents et de l'examinateur. Trois examinateurs, ayant tous reçu une formation sur les procédures et la cotation, ont administré les différentes tâches. Toutes les entrevues ont été filmées, ce qui permettait aux examinateurs de visionner les entrevues à plusieurs reprises, et ainsi, d'assurer d'une codification juste. Pour les enfants de maternelle au DT et ayant un TDL, l'ensemble des tâches a été réalisé en une seule rencontre d'une durée de 1h30, qui incluait une pause au milieu de la tâche de compréhension de phrases. Pour la majorité des enfants d'âge préscolaire, l'administration des tâches s'est déroulée sur deux rencontres de 45 minutes, puisqu'une seule rencontre était trop exigeante pour ces enfants.

Pour la tâche de compréhension de phrases, un item de pratique était d'abord présenté à l'enfant. Lorsque l'enfant était prêt à débuter la tâche, la consigne suivante lui était donnée : « On va regarder des images. Tu pointeras du doigt l'image qui représente chaque phrase que je te dis. Un peu plus loin, tu pourrais ne pas être certain(e) de comprendre la phrase, mais regarde toutes les images comme il faut et montre du doigt celle que tu crois être la bonne ». La consigne était répétée si l'enfant pointait plus d'une image à la fois. L'item pouvait être répété deux fois, au besoin. Seuls des encouragements ne contenant aucun indice étaient donnés par l'examinateur durant la tâche. L'examinateur notait les réponses de l'enfant sur une grille et indiquait si celles-ci étaient bonnes ou mauvaises. Une auto correction spontanée était considérée comme une bonne réponse.

Analyses

Des analyses descriptives (moyenne et écart-type) ont d'abord été utilisées pour décrire les variables socio-démographiques des participants et leurs performances à la tâche de compréhension de phrases. Afin d'observer s'il existait des différences entre les groupes à la tâche de compréhension de phrases, le pourcentage de réponses correctes a été utilisé comme variable dépendante. Comme un pourcentage de réponses correctes constitue plausiblement une échelle ordinale, toutes les analyses de variance à mesures répétées ont été réalisées avec le progiciel *nonparametric analysis of longitudinal data* (nparLD; version 2.1) disponible sur la plateforme statistique R (version 3.3.3). Cependant, le progiciel nparLD ne permet d'inclure qu'au maximum deux variables intra groupe lorsqu'il y a une variable indépendante inter groupe, de sorte que cinq analyses de variance à mesures répétées ont dû être réalisées. D'abord, deux analyses ont été utilisées pour voir les différences entre les groupes (variable indépendante inter groupe) quant au pourcentage de bonnes réponses, soit une pour les catégories de phrases simples (c.-à-d. phrases actives, phrases avec un nom en position objet et questions sujets courtes; variable intra groupe) et une pour les catégories de phrases complexes (c.-à-d. phrases relatives, phrases passives, phrases avec un pronom en position objet et questions complexes; variable intra groupe). De plus, afin d'observer les effets de la complexité et de la longueur des phrases, trois analyses de variance à mesures répétées supplémentaires ont été réalisées. Pour chacune de ces trois analyses, la variable inter groupe (c.-à-d. le groupe) est demeurée constante, alors que la variable ou les variables intra groupe ont variées : (1) la voix (active ou passive), (2) l'agent en position objet (nom ou pronom) et (3) la longueur (questions courtes ou longues) x le déplacement (sans ou avec). Le progiciel nparLD ne procure cependant pas de tests posthoc; des tests de Wilcoxon ont donc été réalisés à part. Lors des tests posthoc, la version rank-sum du test de Wilcoxon a été utilisée pour comparer les groupes et la version signed-rank pour comparer les mesures répétées. Un alpha de 0,05 a été utilisé pour déterminer la signification des différences observées. De plus, bien qu'il s'agisse de mesures répétées et d'analyses non paramétriques, la statistique Δ de Glass a été utilisée pour calculer la taille d'effet. Cette statistique s'interprète comme la statistique *d* de cohen. Afin de présenter certains des résultats, les graphiques de type *Relative treatment effects* (RTE) et des diagrammes en violons ont été utilisés. À noter que le progiciel nparLD est la seule analyse de variance qui accepte les changements de forme de distribution entre les conditions. Ainsi, les diagrammes en violons permettent d'illustrer les changements de forme, et donc, de mieux rendre compte des résultats.

Résultats

Performance des enfants en compréhension de phrases

Les résultats à la tâche de compréhension de phrases sont présentés dans le tableau 3. Pour la compréhension des phrases simples, l'analyse révèle un effet de groupe significatif (p = 0,002), mais pas d'effet de la catégorie de phrase (p = 0,09). Les analyses post-hoc révèlent que les enfants ayant un TDL obtiennent des performances significativement plus faibles que les enfants du groupe DT-maternelle aux phrases actives (p = 0,006, Δ = 1,01) et aux phrases avec un nom en position objet (p = 0,045, Δ = 0,78). Aucune différence significative n'est observée entre le groupe d'enfants ayant un TDL et les enfants du groupe DT-préscolaire. Les enfants du groupe DT-préscolaire

obtiennent également des résultats significativement inférieurs comparativement aux enfants du groupe DT-maternelle pour les phrases actives ($p < 0{,}001$, $\Delta = 1{,}29$) et les phrases avec un nom en position objet ($p = 0{,}029$, $\Delta = 1{,}34$). Quant aux questions sujets courtes, il n'y a pas de différence significative entre les trois groupes. Pour la compréhension des phrases complexes, des effets de groupe ($p = 0{,}0001$) et de catégorie ($p < 0{,}0001$) significatifs sont observés. Les analyses post-hoc révèlent que les enfants ayant un TDL obtiennent des performances significativement plus faibles que les enfants du groupe DT-maternelle aux phrases passives ($p < 0{,}001$, $\Delta = 1{,}68$), aux phrases avec un pronom en position objet ($p < 0{,}001$, $\Delta = 1{,}41$) et aux questions complexes ($p = 0{,}026$, $\Delta = 0{,}89$). Toutefois, aucune différence significative n'est observée entre les groupes en ce qui à trait à la compréhension des phrases relatives ($p = 0{,}221$). Les enfants ayant un TDL obtiennent des résultats qui ne diffèrent pas significativement des enfants du groupe DT-préscolaire, et ce, pour toutes les catégories de phrase. Les enfants du groupe DT-préscolaire obtiennent également des résultats significativement inférieurs comparativement aux enfants du groupe DT-maternelle pour les phrases passives ($p < 0{,}001$, $\Delta = 1{,}27$), les phrases avec un pronom en position objet ($p = 0{,}01$, $\Delta = 1{,}00$) et les questions ($p < 0{,}001$, $\Delta = 1{,}61$). Le tableau 3 décrit les performances aux différents types de phrases pour les trois groupes d'enfants à l'étude.

Effet de la complexité et de la longueur

Effet de la voix passive. Un effet significatif de la voix passive sur les performances des trois groupes ($p < 0{,}001$) a été observé. En effet, les performances aux phrases passives sont significativement inférieures aux performances aux phrases actives pour les enfants ayant un TDL ($p = 0{,}014$, $\Delta = 1{,}10$), les enfants du groupe DT-maternelle ($p = 0{,}013$, $\Delta = 0{,}73$) et les enfants du groupe DT-préscolaire ($p = 0{,}002$, $\Delta = 0{,}67$). La figure 1 montre le diagramme en violon de la distribution et la différence entre les RTE pour les phrases passives et actives, et ce, pour chacun des groupes.

Effet de la catégorie grammaticale en position objet. Un effet de la catégorie grammaticale en position objet est observé pour tous les groupes ($p = 0{,}002$). Les performances pour les phrases avec un pronom en position objet sont significativement plus faibles que les performances aux phrases avec un nom en position objet, et ce, tant pour les enfants ayant un TDL ($p < 0{,}001$, $\Delta = 2{,}22$) que les enfants du groupe DT-maternelle ($p < 0{,}001$, $\Delta = 1{,}66$) et les enfants du groupe DT-préscolaire ($p < 0{,}001$, $\Delta = 0{,}83$). La figure 2 montre le diagramme en violon de la distribution et la différence entre les RTE aux phrases avec un pronom en position objet pour chacun des groupes.

Effet du déplacement de syntagme et de la longueur des questions. Les analyses ne révèlent pas d'effet du déplacement de syntagme dans la performance aux questions avec un pronom interrogatif « qui » ($p = 0{,}12$), et ce, peu importe le groupe. Des effets de groupe ($p = 0{,}002$) et de longueur ($p < 0{,}001$) sont toutefois observés. Les analyses post-hoc révèlent que la longueur a un effet significatif sur les performances pour les enfants de maternelle au DT ($p = 0{,}001$, $\Delta = 0{,}53$ pour les questions sujets; $p = 0{,}004$, $\Delta = 0{,}23$ pour les questions objets) et pour les enfants d'âge préscolaire au DT ($p < 0{,}001$, $\Delta = 1{,}01$ pour les questions sujets; $p = 0{,}051$, $\Delta = 0{,}46$ pour les questions objets). Toutefois, bien que les enfants ayant un TDL obtiennent des performances plus faibles aux questions longues qu'aux questions courtes, les différences ne rejoignent pas un seuil statistique significatif ($p = 0{,}10$ pour les questions sujets; $p = 0{,}21$ pour les questions objets). Les analyses révèlent également un effet combiné de la longueur et du déplacement ($p = 0{,}043$). Ainsi, une différence significative est observée entre les performances aux questions sujets courtes et les questions objets longues, et ce, tant pour les enfants de maternelle ayant un TDL ($p = 0{,}020$, $\Delta = 0{,}62$) que pour les enfants de maternelle au DT ($p = 0{,}001$, $\Delta = 0{,}89$) et les enfants d'âge préscolaire au DT ($p = 0{,}003$, $\Delta = 0{,}88$). La figure 3 représente le diagramme en violon de la distribution et le RTE aux questions courtes et longues en fonction du groupe et de la présence d'un déplacement de syntagme.

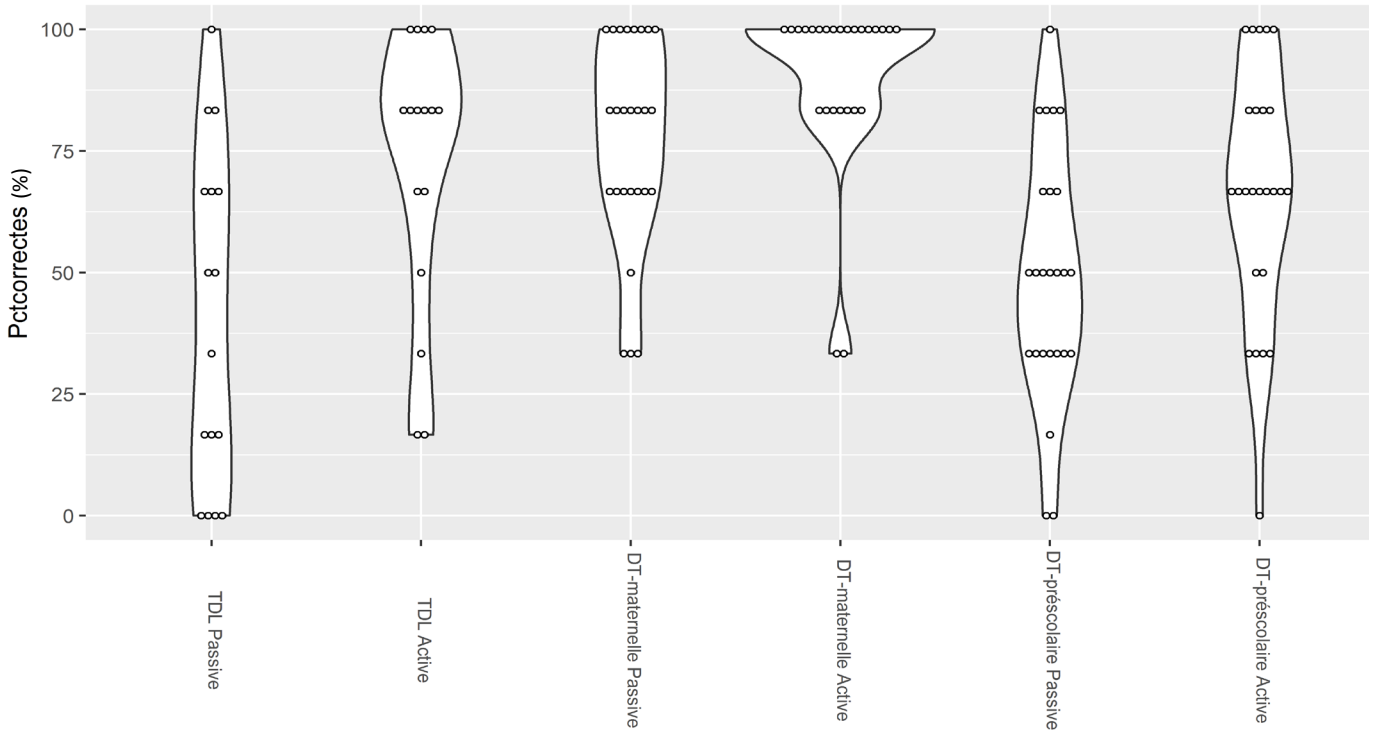
Discussion

Le but de cette étude était d'explorer les habiletés de compréhension de phrases simples et complexes des enfants francophones de maternelle ayant un TDL. Les résultats montrent qu'il existe des différences entre les performances des enfants ayant un TDL et les enfants au DT sur le plan de la compréhension syntaxique. L'étude permet également de corroborer les études antérieures en montrant l'effet de la complexité syntaxique et de la longueur sur la compréhension des enfants ayant un TDL. Aussi, les performances pour les différents groupes permettent de dresser un portrait des habiletés en compréhension syntaxique d'enfants francophones âgés entre 4 et 5 ans.

Tableau 3. Pourcentage moyen de réponses adéquates aux différents types de phrases pour les trois groupes à l'étude

Types de phrases	TDL (*n* = 16)	DT-maternelle (*n* = 26)	DT-préscolaire (*n* = 25)
Simples			
Actives (*n* = 6)			
Moyenne (ÉT)	71,88 (28,36)	90,38 (18,36)*	66,67 (25,91)
Étendue	16,7-100	33,33-100	0-100
Nom en position objet (*n* = 8)			
Moyenne (ÉT)	81,25 (13,70)	89,42 (10,42)*	75,5 (22,38)
Étendue	50-100	62,5-100	25-100
Questions sujets courtes (*n* = 6)			
Moyenne (ÉT)	72,92 (23,47)	82,69 (17,31)	71,33 (25,69)
Étendue	16,7-100	50-100	0-100
Complexes			
Relatives (*n* = 4)			
Moyenne (ÉT)	45,31 (29,18)	60,58 (23,64)	56,00 (28,21)
Étendue	0-100	0-100	0-100
Passives (*n* = 6)			
Moyenne (ÉT)	40,63 (34,41)	76,92 (21,64)*	49,33 (25,68)
Étendue	0-100	33,33-100	0-100
Pronom en position objet (*n* = 8)			
Moyenne (ÉT)	50,78 (20,14)	72,11 (15,13)*	57,00 (21,07)
Étendue	25-87,50	37,5-100	25-100
Questions complexes (*n* = 18)			
Moyenne (ÉT)	60,76 (15,51)	72,22 (12,86)*	51,56 (19,68)
Étendue	33,33-88,89	50-94,44	11,11-88,89

Note. *Différences significatives avec les deux autres groupes ($p < 0,05$). ÉT = écart-type; DT-maternelle = groupe d'enfants au développement typique appariés sur l'âge; DT-préscolaire = groupe d'enfants au développement typique appariés sur les habiletés de vocabulaire réceptif; TDL = groupe d'enfants ayant un trouble développemental du langage.



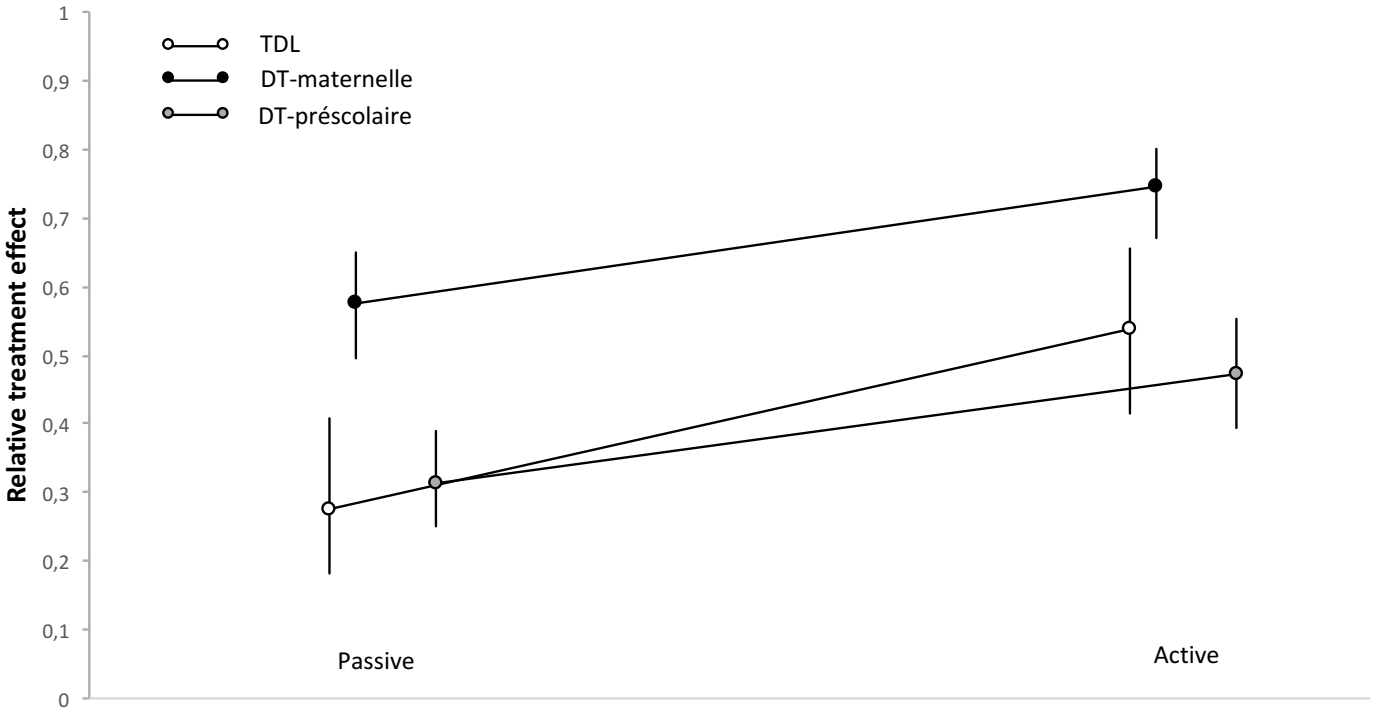


Figure 1. Diagramme en violon et intervalle de confiance à 95% pour les performances aux phrases aux voix passive et active. Active = phrases à la voix active; DT-maternelle = groupe d'enfants au développement typique appariés sur l'âge; DT-préscolaire = groupe d'enfants au développement typique appariés sur les habiletés de vocabulaire réceptif; Passive = phrases à la voix passive; PctCorrectes = pourcentage de réponses correctes; TDL = groupe d'enfants ayant trouble développemental du langage.

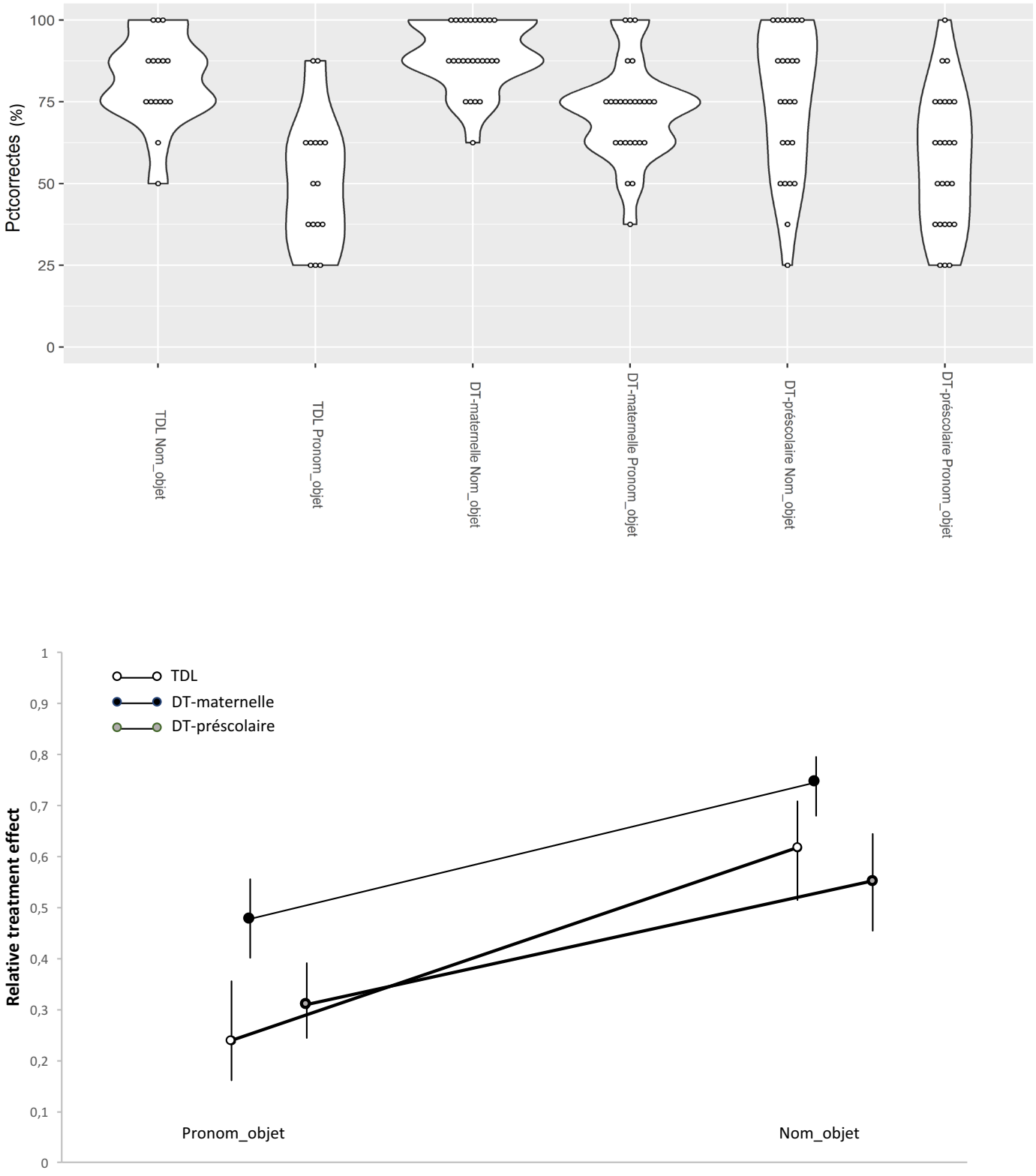


Figure 2. Diagramme en violon et intervalle de confiance à 95% pour les performances aux phrases avec un nom ou un pronom en position objet. DT-maternelle = groupe d'enfants au développement typique appariés sur l'âge; DT-préscolaire = groupe d'enfants au développement typique appariés sur les habiletés de vocabulaire réceptif; Nom_objet = phrases avec un nom en position objet; Pronom_objet = phrases avec un pronom en position objet; PctCorrectes = pourcentage de réponses correctes; TDL = groupe d'enfants ayant un trouble développemental du langage.

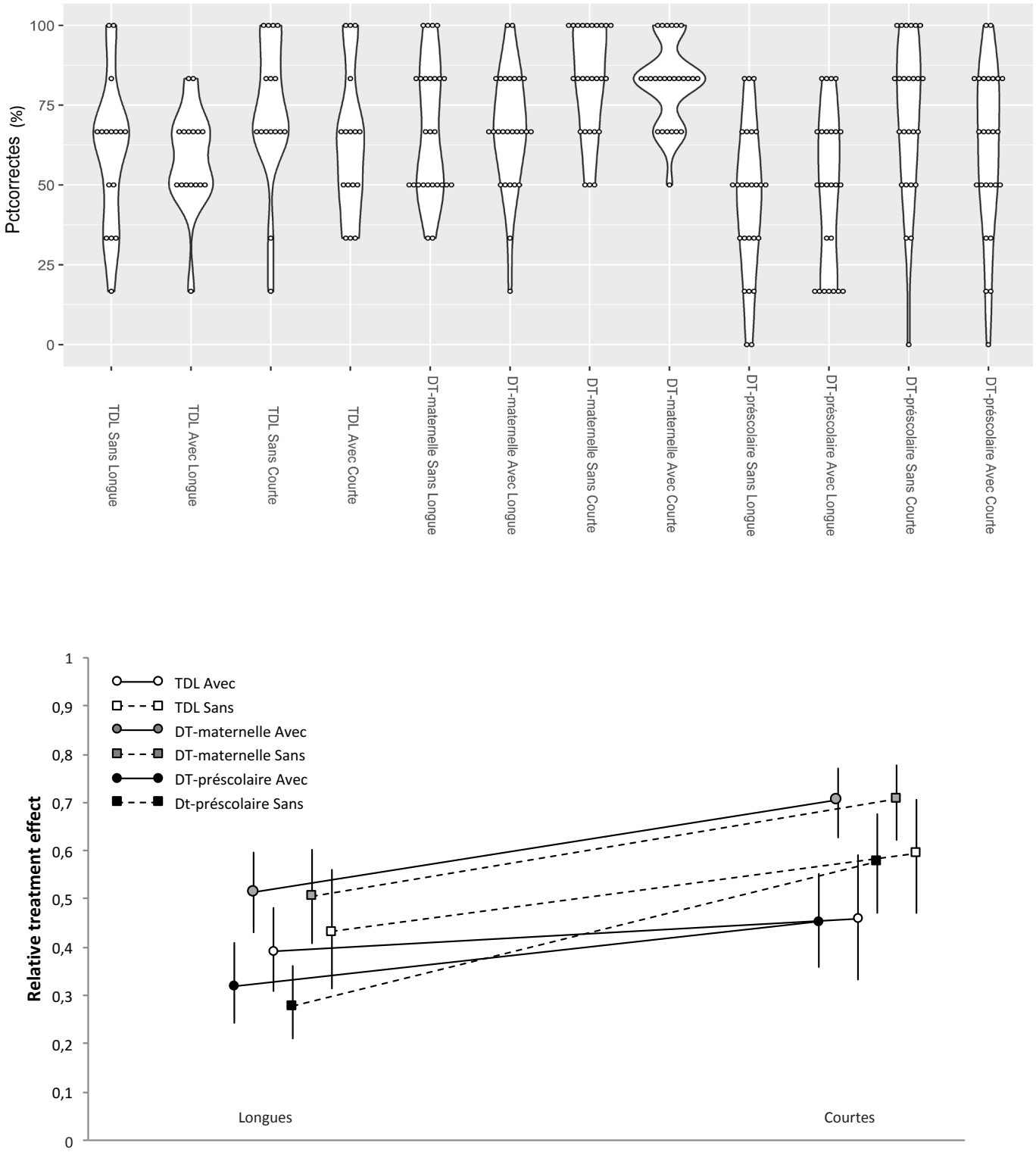


Figure 3. Diagramme en violon et intervalle de confiance à 95% pour les performances aux questions courtes et longues en fonction de la présence ou de l'absence d'un déplacement de syntagme. Avec = questions avec un déplacement de syntagme; Courtes = questions courtes; DT-maternelle = groupe d'enfants au développement typique appariés sur l'âge; DT-préscolaire = groupe d'enfants au développement typique appariés sur les habiletés de vocabulaire réceptif; Longues = questions longues; PctCorrectes = pourcentage de réponses correctes; Sans = questions sans déplacement de syntagme; TDL = groupe d'enfants ayant un trouble développemental du langage.

Habiletés de compréhension de phrases simples et complexes des enfants ayant un TDL

Phrases simples. Les résultats observés à la compréhension des phrases simples (c.-à-d. phrases actives, phrases avec un nom en position objet et questions sujets courtes) contredisent certaines études effectuées auprès d'enfants anglophones qui rapportaient une performance similaire aux phrases simples entre les enfants ayant un TDL et les enfants au DT du même âge (Montgomery, 2004; Montgomery et Evans, 2009). Dans la présente étude, les résultats montrent que les enfants ayant un TDL performent significativement moins bien que les enfants au DT du même âge. Leurs performances aux questions simples correspondent plutôt aux performances observées chez des enfants plus jeunes. Ainsi, la compréhension syntaxique de phrases simples ne serait pas pleinement acquise par tous les enfants âgés de 5 ans ayant un TDL. Les résultats indiquent également une grande variabilité entre les performances des différents enfants. En effet, la moyenne des performances des enfants peut varier entre 0% et 100% selon le type de phrases simples. Ces résultats pourraient toutefois être expliqués par des différences entre l'échantillon de la présente étude et ceux des études rapportées. En effet, les études portant sur le développement syntaxique des enfants anglophones ont étudié la compréhension de phrases simples chez des enfants âgés entre 6 et 9 ans, et donc, plus âgés que les participants de la présente étude.

Phrases complexes. Les enfants ayant un TDL auraient plus de difficultés à comprendre différentes phrases complexes que les enfants au DT du même âge, ce qui confirme les résultats d'études antérieures portant sur le développement syntaxique des enfants anglophones (Bishop et al., 2000; Deevy et Leonard, 2004; Leonard et al., 2013; Montgomery, 2000a, 2000b, 2004; Montgomery et Evans, 2009). La présente étude vient également appuyer la majorité des catégories de phrases reconnues comme étant plus difficiles à comprendre par les enfants anglophones ayant un TDL selon Montgomery et ses collaborateurs (2016). En effet, les résultats de la présente étude montrent que les enfants ayant un TDL ont plus de difficultés que les enfants au DT du même âge à comprendre les phrases passives, les phrases avec un pronom en position objet et les questions complexes avec un pronom interrogatif « qui » (c.-à-d. les questions objets et les questions longues). Leur performance serait plutôt comparable à celle des enfants d'âge préscolaire au DT, ce qui corrobore les études qui avaient comparé des enfants ayant un TDL à des enfants plus jeunes appariés sur le langage (Deevy et Leonard, 2004; Leonard, 2014a; Montgomery, 2000a, 2000b, 2004; Montgomery et Evans, 2009). Selon ces études, les enfants ayant un TDL avaient des performances comparables, voire plus faibles que les enfants plus jeunes.

Bien que les performances des enfants ayant un TDL soient inférieures à celles des enfants au DT pour la compréhension des phrases relatives, la différence ne s'avère pas significative. Ainsi, ce résultat contredit l'idée que les phrases relatives seraient plus difficiles à comprendre pour les enfants ayant un TDL que pour les enfants au DT (Montgomery et al., 2016). Toutefois, ce résultat peut être expliqué par l'âge des participants de la présente étude. D'abord, il est attendu que les phrases relatives soient comprises entre l'âge de 4 et 7 ans (de Villiers et al., 1979). Il est donc possible que la compréhension de ces phrases soit encore en développement, et ce, chez tous les enfants à l'étude. De plus, l'étude de Montgomery (2004) a montré des différences entre les enfants ayant un TDL et les enfants au DT pour la compréhension des phrases relatives entre l'âge de 6 et 10 ans. Ainsi, il est possible que les difficultés en lien avec la compréhension des phrases relatives objets soient davantage marquées chez des enfants plus vieux que ceux de la présente étude. Il est aussi important de souligner que Montgomery et ses collaborateurs (2016) ont rapporté que les enfants ayant un TDL présentaient davantage de difficultés dans la compréhension des phrases relatives objets. Toutefois, dans la présente étude, les phrases relatives n'étaient pas toutes des phrases relatives objets, bien qu'elles comprenaient toutes un déplacement syntaxique. Il est donc possible que la nature de la phrase relative ait influencé les résultats.

Effet de la complexité et de la longueur sur la compréhension syntaxique. À l'instar des études portant sur les enfants anglophones, les résultats de cette étude mettent en évidence les effets de la complexité et de la longueur des phrases sur la compréhension de phrases des enfants francophones. Tel que soulevé dans l'étude de Montgomery (2004), la voix passive est plus difficilement comprise que la voix active par les enfants ayant un TDL, et ce, en raison de la forme non canonique de ces phrases. La présente étude corrobore également les résultats de Bishop et ses collaborateurs (2000) et de Montgomery et Evans (2009) quant aux difficultés en lien avec la catégorie grammaticale de l'objet dans la phrase des enfants ayant un TDL.

De plus, comme dans l'étude de Deevy et Leonard (2004), nos résultats confirment un effet combiné de la longueur et du déplacement pour la compréhension des questions objets longues. Toutefois, pris indépendamment, seule la longueur a un effet significatif sur les performances, ce qui corrobore les résultats de plusieurs études (Deevy

et Leonard, 2004; Montgomery, 2004; Norbury, Bishop et Briscoe, 2002). Ainsi, contrairement à une hypothèse mise de l'avant par van der Lely (van der Lely, 2004, 2005; van der Lely et al., 1998), le mouvement syntaxique dans les questions objets ne serait pas, à lui seul, un élément discriminant de la performance en compréhension syntaxique entre les enfants francophones ayant un TDL et les enfants francophones au DT du même âge. Nos résultats sont donc mieux expliqués par des hypothèses basées sur des processus mnésiques. En effet, selon certains auteurs, les déficits mnésiques des enfants ayant un TDL pourraient être suffisamment importants pour expliquer les problèmes langagiers (van Daal, Verhoeven et van Balkom, 2009; Vugs, Knoors, Cuperus, Hendriks et Verhoeven, 2016). Différentes études ont par ailleurs tenté d'établir des corrélations entre les performances en compréhension de phrases et les performances en mémoire via des tâches de répétition de non-mots (Fortunato-Tavares et al., 2012; Marton et Schwartz, 2003; Montgomery, 1995, 2000a, 2000b; Montgomery et al., 2008; van Daal et al., 2009; Vugs et al., 2016). Les hypothèses basées sur des processus mnésiques proposent que les enfants ayant un TDL auraient une plus faible capacité de mémoire verbale à court-terme, soit de moindres habiletés à maintenir une information active durant une activité cognitive (Montgomery et al., 2016). Ainsi, les phrases complexes de structure non canonique nécessiteraient un maintien des informations plus longtemps en mémoire que les phrases canoniques qui ne requièrent qu'un traitement de gauche à droite (Montgomery et al., 2016). Toutefois, étant donné les résultats contradictoires observés dans la littérature, de plus amples études sont nécessaires pour comprendre de façon exhaustive les difficultés syntaxiques des enfants ayant un TDL.

Développement typique de la compréhension syntaxique. En plus de donner un portrait des éléments difficiles chez les enfants ayant un TDL, cette étude apporte des clarifications quant au développement syntaxique réceptif des enfants francophones au DT âgés entre 4 et 5 ans. D'abord, les résultats indiquent que peu de structures syntaxiques sont pleinement acquises au plan réceptif par les enfants de 4 ans. En effet, bien que la littérature indique que les enfants comprennent les phrases actives entre l'âge de 3 et 4 ans (Evans, 2002), les résultats de cette étude tendent à montrer que les phrases actives (et les phrases simples de façon générale) pourraient n'être encore que partiellement comprises par certains enfants âgés de 4 ans. Les performances montrent également une grande variabilité inter sujets. En effet, les performances aux phrases simples chez les enfants de 4 ans varient entre 0% et 100%. Ainsi, le développement de la compréhension syntaxique semble se poursuivre à cet âge. Quant aux enfants francophones de 5 ans, la plupart ont acquis une bonne compréhension des phrases simples. Une grande variabilité inter individuelle est néanmoins encore présente pour les phrases complexes à cet âge. Les résultats quant aux phrases complexes corroborent les données concernant le développement normal de la compréhension syntaxique des enfants anglophones. En effet, les études précédemment publiées ont montré que la compréhension des phrases passives et des phrases avec pronom en position objet émergerait entre l'âge de 5 et 6 ans (Montgomery et al., 2008; Paul et Norbury, 2012), alors que les phrases relatives seraient comprises entre l'âge de 4 et 7 ans (de Villiers et al., 1979). Les difficultés observées quant à la compréhension des phrases complexes chez les enfants ayant un TDL, mais également chez les enfants au DT âgés entre 4 et 5 ans peuvent être expliquées par l'effet de la complexité et de la longueur des phrases. En effet, les résultats montrent que tous les enfants à l'étude éprouvent de la difficulté à comprendre les phrases non canoniques, soit les phrases passives et les phrases avec un pronom en position objet. La moyenne des performances en compréhension aux phrases relatives montre également une faible maîtrise de ce type de phrase, et ce, chez tous les enfants au DT. Ainsi, les enfants au DT âgés entre 4 et 5 ans seraient encore sensibles à la complexité et à la longueur des phrases. Finalement, les résultats permettent de confirmer que les enfants au DT ne sont pas influencés par le déplacement de syntagme dans la formulation d'une question et acquerraient la compréhension des questions sujets et objets en même temps (Stromswold, 1995). En somme, cette étude montre que plusieurs éléments de compréhension syntaxique sont encore en développement entre l'âge de 4 et 5 ans, résultats qui convergent avec ceux de l'étude de Delage et Frauenfelder (2012) qui observaient un développement de cette composante jusqu'à l'âge de 12 ans.

Certaines forces et faiblesses de l'étude doivent être mises en relief. D'abord, la principale force de cette étude est d'apporter de l'information quant à la compréhension syntaxique des enfants francophones de maternelle au DT et ayant un TDL, un sujet peu, voire pas du tout, exploré dans la littérature. Ces informations pourront ainsi être utiles aux cliniciens œuvrant avec cette clientèle afin d'ajuster leurs méthodes d'évaluation et d'intervention. Lors de futures recherches, il pourra être pertinent d'approfondir certains éléments de la compréhension syntaxique en ciblant de façon plus spécifique les différentes catégories de phrase. De plus, la formation de deux groupes contrôle, l'un apparié en fonction de l'âge, l'autre en fonction du vocabulaire réceptif, est une force de cette étude. D'une part, le groupe d'enfants au DT appariés sur l'âge permet de

mieux objectiver les difficultés qui sont présentes chez les enfants ayant un TDL et qui ne sont pas dues à l'âge précoce des enfants. D'autre part, le groupe d'enfants appariés en fonction du vocabulaire réceptif permet de voir que la performance des enfants présentant un TDL correspond davantage à celles des enfants plus jeunes. Cela corrobore d'ailleurs la proposition que les enfants ayant un TDL auraient un développement de l'acquisition des habiletés syntaxiques plus lent que celui des enfants au DT et non un développement atypique (Leonard, 2014b). Les hypothèses explicatives, dont les liens entre la compréhension syntaxique et les difficultés mnésiques, sont une avenue peu explorée en contexte franco-québécois et qui aurait avantage à être considérée pour de futures pistes de recherche.

L'étude comporte également certaines faiblesses. D'abord, il importe d'apporter certaines nuances aux résultats. Les performances lors d'une tâche ciblée de compréhension syntaxique peuvent ne pas refléter le fonctionnement au quotidien des enfants au plan réceptif. En effet, lors d'une question, consigne ou énoncé donné en contexte, les enfants peuvent se fier aux expressions faciales, à l'intonation ou aux objets de leur environnement, ce qui facilite leur compréhension (Paul et Norbury, 2012). Ainsi, leurs performances lors d'une tâche décontextualisée, telle qu'utilisée dans le cadre de cette étude, sont inférieures à ce que l'enfant peut comprendre lorsqu'il s'appuie sur le contexte. Malgré cela, il est intéressant de connaître ce que ces enfants comprennent réellement en dehors du contexte afin de soutenir leur développement à long terme, tant au plan réceptif qu'expressif. En effet, une meilleure connaissance des formes de phrases comprises sur le plan linguistique permettra aux intervenants de mieux cibler les objectifs d'intervention (Paul et Norbury, 2012). De plus, certaines catégories de phrases contiennent un nombre limité de phrases (p. ex. les phrases relatives). Un plus grand nombre de phrases dans chacune des catégories aurait potentiellement permis un portrait plus complet des habiletés syntaxiques. Toutefois, considérant l'âge des enfants impliqués dans cette étude, il était important de considérer la longueur de la tâche afin d'avoir une pleine collaboration des enfants, et ainsi, des résultats plus justes vis-à-vis leur potentiel. Finalement, les caractéristiques lexicales (âge d'acquisition, fréquence, etc.) des mots utilisés dans les phrases n'ont pas été prises en considération dans le cadre de cette étude, ce qui serait un ajout intéressant pour une future étude.

Perspectives cliniques

Nos résultats ont des retombées pour la pratique clinique, et ce, tant sur le plan de l'évaluation que de l'intervention. D'une part, ce portrait plus détaillé du développement syntaxique typique des enfants francophones âgés entre 4 et 5 ans permet de mieux mesurer l'ampleur des déficits chez les enfants en difficulté lors de l'évaluation. Les tâches utilisées dans notre étude peuvent aussi servir d'inspiration pour des outils à utiliser en clinique. D'autre part, une meilleure connaissance de la compréhension syntaxique des enfants ayant un TDL permettra aux professionnels œuvrant auprès de ces enfants d'ajuster leurs objectifs d'intervention. Par ailleurs, en plus de contribuer aux connaissances quant à la compréhension syntaxique des enfants franco-québécois avec ou sans difficultés de langage, les résultats ont une portée pour le développement langagier des enfants francophones en général. En effet, la syntaxe d'une même langue étant similaire d'une culture à l'autre, les éléments difficiles syntaxiquement dans une population d'enfants francophones sont potentiellement difficiles pour tous les enfants francophones.

Conclusion

Cette étude contribue à une meilleure description des habiletés de compréhension syntaxique des enfants francophones de maternelle ayant un TDL, ainsi que de celles des enfants francophones d'âge préscolaire et de maternelle au DT. Les résultats montrent que les enfants de maternelle ayant un TDL éprouvent des difficultés de compréhension syntaxique en lien avec la complexité et la longueur des phrases. De plus, les résultats invitent les professionnels à prendre en compte ces difficultés en ajustant leurs évaluations et interventions aux besoins de l'enfant, tout en ayant une meilleure connaissance du développement réceptif syntaxique typique.

Références

Bishop, D. V. M., Bright, P., James, C., Bishop, S. J. et van der Lely, H. K. J. (2000). Grammatical SLI: A distinct subtype of developmental language impairment? *Applied Psycholinguistics, 21*, 159-181.

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T. et CATALISE consortium. (2016). CATALISE: A multinational and multidisciplinary delphi consensus study. Identifying language impairments in children. *PLoS One, 11*(12), 1-26. doi:10.1371/journal.pone.0158753.

Botting, N., Durkin, K., Toseeb, U., Pickles, A. et Conti-Ramsden, G. (2016). Emotional health, support, and self-efficacy in young adults with a history of language impairment. *British Journal of Developmental Psychology, 34*, 538-554. doi:10.1111/bjdp.12148

Conti-Ramsden, G., Durkin, K., Simkin, Z. et Knox, E. (2009). Specific language impairment and school outcomes. I: Identifying and explaining variability at the end of compulsory education. *International Journal of Language & Communication Disorders, 44*, 15-35. doi:10.1080/13682820801921601

De Villiers, J. G., Tager Flusberg, H. B., Hakuta, K. et Cohen, M. (1979). Children's comprehension of relative clauses. *Journal of Psycholinguistic Research, 8*, 499-518. doi: 10.1007/BF01067332

Deevy, P. et Leonard, L. B. (2004). The comprehension of wh-questions in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 47*, 802-815. doi:10.1044/1092-4388(2004/060)

Delage, H. et Frauenfelder, U. (2012). Développement de la mémoire de travail et traitement des phrases complexes: Quelle relation? *SHS Web of Conferences, 1*, 1555-1573. doi:10.1051/shsconf/20120100141

Dunn, L. M., Thériault-Whalen, C. M. et Dunn, L. M. (1993). *Échelle de vocabulaire en images Peabody. Adaptation française du Peabody Picture Vocabulary Test*. Toronto, Canada : Psycan.

Ebbels, S. (2007). Teaching grammar to school-aged children with specific language impairment using shape coding. *Child Language Teaching and Therapy, 23*, 67-93. doi:10.1191/0265659007072143

Evans, J. L. (2002). Variability in comprehension strategy use in children with SLI: A dynamical systems account. *International Journal of Language & Communication Disorders, 37*, 95-116. doi:10.1080/13682820110116767

Fortunato-Tavares, T., de Andrade, C. R. F., Befi-Lopes, D. M., Hestvik, A., Epstein, B., Tornyova, L. et Schwartz, R. G. (2012). Syntactic structural assignment in Brazilian Portuguese-speaking children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 55*, 1097-1111. doi:10.1044/1092-4388(2011/10-0215)

Johnson, C. J., Beitchman, J. H. et Brownlie, E. B. (2010). Twenty-year follow-up of children with and without speech-language impairments: Family, educational, occupational, and quality of life outcomes. *American Journal of Speech-Language Pathology, 19*, 51-65. doi:10.1044/1058-0360(2009/08-0083)

Justice, L. M., Bowles, R. P., Pence Turnbull, K. L. et Skibbe, L. E. (2009). School readiness among children with varying histories of language difficulties. *Developmental Psychology, 45*, 460-476. doi:10.1037/a0014324

Leclercq, A.-L., Majerus, S., Prigent, G. et Maillart, C. (2013). The impact of dual tasking on sentence comprehension in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 56*, 265-280. doi:10.1044/1092-4388(2012/10-0290)

Leonard, L. B. (2014a). *Children with specific language impairment* (2e éd.). Cambridge, MA : MIT Press.

Leonard, L. B. (2014b). Specific language impairment across languages. *Child Development Perspectives, 8*, 1-5. doi: 10.1111/cdep.12053

Leonard, L. B., Deevy, P., Fey, M. E. et Bredin-Oja, S. L. (2013). Sentence comprehension in specific language impairment: A task designed to distinguish between cognitive capacity and syntactic complexity. *Journal of Speech, Language, and Hearing Research, 56*, 577-589. doi:10.1044/1092-4388(2012/11-0254)

Lin, Y.-A. (2006). Against the deficit in computational grammatical complexity hypothesis: A corpus-based study. *Concentric: Studies in Linguistics, 32*(2), 59-7

Marinis, T. et van der Lely, H. K. J. (2007). On-line processing of wh-questions in children with G-SLI and typically developing children. *International Journal of Language & Communication Disorders, 42*, 557-582. doi:10.1080/13682820601058190

Marton, K. et Schwartz, R. G. (2003). Working memory capacity and language processes in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 46*, 1138-1153. doi:10.1044/1092-4388(2003/089)

Montgomery, J. W. (1995). Sentence comprehension in children with specific language impairment: The role of phonological working memory. *Journal of Speech, Language, and Hearing Research, 38*, 187-199. doi:10.1044/jshr.3801.187

Montgomery, J. W. (2000a). Relation of working memory to off-line and real-time sentence processing in children with specific language impairment. *Applied Psycholinguistics, 21*, 117-148.

Montgomery, J. W. (2000b). Verbal working memory and sentence comprehension in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 43*, 293-308. doi:10.1044/jslhr.4302.293

Montgomery, J. W. (2003). Working memory and comprehension in children with specific language impairment: What we know so far. *Journal of Communication Disorders, 36*, 221-231. doi:10.1016/S0021-9924(03)00021-2

Montgomery, J. W. (2004). Sentence comprehension in children with specific language impairment: Effects of input rate and phonological working memory. *International Journal of Language & Communication Disorders, 39*, 115-133. doi:10.1080/13682820310001616985

Montgomery, J. W. et Evans, J. L. (2009). Complex sentence comprehension and working memory in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 52*, 269-288. doi:10.1044/1092-4388(2008/07-0116)

Montgomery, J. W., Gillam, R. B. et Evans, J. L. (2016). Syntactic versus memory accounts of the sentence comprehension deficits of specific language impairment: Looking back, looking ahead. *Journal of Speech, Language, and Hearing Research, 59*, 1491-1504. doi:10.1044/2016_JSLHR-L-15-0325

Montgomery, J. W., Magimairaj, B. M. et O'Malley, M. H. (2008). Role of working memory in typically developing children's complex sentence comprehension. *Journal of Psycholinguistic Research, 37*, 331-354. doi:10.1007/s10936-008-9077-z

Norbury, C. F., Bishop, D. V. M. et Briscoe, J. (2002). Does impaired grammatical comprehension provide evidence for an innate grammar module? *Applied Psycholinguistics, 23*, 247-268. doi:10.1017/S0142716402002059

Paul, R. et Norbury, C. (2012). *Language disorders from infancy through adolescence: Listening, speaking, reading, writing, and communicating* (4e éd.). St-Louis, MO : Elsevier.

Pettenati, P., Benassi, E., Deevy, P., Leonard, L. B. et Caselli, M. C. (2015). Extra-linguistic influences on sentence comprehension in Italian-speaking children with and without specific language impairment. *International Journal of Language & Communication Disorders, 50*, 312-321. doi:10.1111/1460-6984.12134

Reilly, S., Tomblin, B., Law, J., McKean, C., Mensah, F. K., Morgan, A., . . . Wake, M. (2014). Specific language impairment: A convenient label for whom? *International Journal of Language & Communication Disorders, 49*, 416-451. doi:10.1111/1460-6984.12102

Segui, J. et Léveillé, M. (1977). Étude sur la compréhension de phrases chez l'enfant. *Enfance, 30*, 105-115. doi:10.3406/enfan.1977.2637

Stromswold, K. (1995). The acquisition of subject and object wh-questions. *Language Acquisition, 4*, 5-48. doi:10.1080/10489223.1995.9671658

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E. et O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research, 40*, 1245-1260. doi:10.1044/jslhr.4006.1245

van Daal, J., Verhoeven, L. et van Balkom, H. (2009). Cognitive predictors of language development in children with specific language impairment (SLI). *International Journal of Language & Communication Disorders, 44*, 639-655. doi:10.1080/13682820802276930

van der Lely, H. K. J. (2004). Evidence for and implications of a domain-specific grammatical deficit. Dans L. Jenkins (dir.), *The genetics of language* (p. 117-145). Oxford, MS : Elsevier.

van der Lely, H. K. J. (2005). Domain-specific cognitive systems: Insight from Grammatical-SLI. *Trends in Cognitive Sciences, 9*, 53-59. doi:10.1016/j.tics.2004.12.002

van der Lely, H. K. J., Rosen, S. et McClelland, A. (1998). Evidence for a grammar-specific deficit in children. *Current Biology, 8*, 1253-1258. doi:10.1016/S0960-9822(07)00534-9

Vugs, B., Knoors, H., Cuperus, J., Hendriks, M. et Verhoeven, L. (2016). Interactions between working memory and language in young children with specific language impairment (SLI). *Child Neuropsychology, 22*, 955-978. doi:10.1080/09297049.2015.105

Remerciements

Cette recherche a été supportée par une subvention du *Fonds québécois de recherche sur la société et la culture* (2008-2011) accordée à Chantal Desmarais. De plus, nous aimerions remercier le Centre interdisciplinaire de recherche en réadaptation et intégration sociale, les assistants de recherche, M. Jean Leblond, statisticien, ainsi que les enfants et leur famille qui ont participé à cette étude.

Note des auteurs

Les demandes au sujet de cet article doivent être adressées à Chantal Desmarais, 1050 avenue de la Médecine, Québec, QC, G1V 0A6.
Courriel : chantal.desmarais@rea.ulaval.ca

Annexe

Liste des phrases utilisées

Essai

La pomme rouge est sur la chaise.

Phrases simples

Actives

1. L'homme poursuit le chien.
2. Le cheval poursuit la fille.
3. Le garçon poursuit le mouton.
4. Le garçon pousse le camion.
5. La vache pousse la dame.
6. La fille pousse le cheval.

Nom en position objet

1. La fille laisse tomber la tasse.
2. Le monsieur arrose les fleurs.
3. L'éléphant porte les enfants.
4. Le cheval regarde la fille.
5. La vache regarde les enfants.
6. Le chat et le chien poursuivent le mouton.
7. La dame porte le bébé.
8. Le garçon regarde les belles fleurs.

Questions sujets courtes

1. Qui est en train de regarder le cheval?
2. Qui est en train de poursuivre l'oiseau?
3. Qui est en train de pousser le zèbre?
4. Qui est en train de laver le chien?
5. Qui est en train d'embrasser le crocodile?
6. Qui est en train de mordre le chat?

Phrases complexes

Relatives

1. Le garçon mange les pommes que la fille cueille.
2. Le monsieur regarde la vache que poursuit le chat.
3. La pomme que mange le garçon est verte.
4. La vache que le chien poursuit est brune.

Passives

1. Le mouton est poursuivi par le garçon.
2. La fille est poursuivie par le cheval.
3. Le camion est poussé par le garçon.
4. La dame est poussée par la vache.
5. Le cheval est poussé par la fille.
6. Le chien est poursuivi par l'homme.

Phrases avec pronom en position objet

1. La vache les regarde.
2. La dame le porte.
3. La fille la laisse tomber.
4. Le cheval la regarde.
5. Le chat et le chien le poursuivent.
6. L'éléphant les porte.
7. Le garçon les regarde.
8. Le monsieur les arrose.

Questions sujets longues

1. Qui est en train d'arroser le petit cochon blanc?
2. Qui est en train de caresser le gentil lapin blanc?
3. Qui est en train de frapper le méchant ours noir?
4. Qui est en train de lancer le gros éléphant brun?
5. Qui est en train de nourrir le méchant loup brun?
6. Qui est en train de chatouiller le petit singe vert?

Questions objets courtes

1. Qui l'oiseau est-il en train de poursuivre?
2. Qui le zèbre est-il en train de pousser?
3. Qui le chat est-il en train de mordre?
4. Qui le crocodile est-il en train d'embrasser?
5. Qui le chien est-il en train de laver?
6. Qui le cheval est-il en train de regarder?

Questions objets longues

1. Qui le petit cochon blanc est-il en train d'arroser?
2. Qui le méchant loup brun est-il en train de nourrir?
3. Qui le petit singe vert est-il en train de chatouiller?
4. Qui le gentil lapin blanc est-il en train de caresser?
5. Qui le méchant ours noir est-il en train de frapper?
6. Qui le gros éléphant brun est-il en train de lancer?

Effectiveness of Two Methods for Teaching Critical Thinking to Communication Sciences and Disorders Undergraduates

Efficacité de deux méthodes d'enseignement pour développer l'esprit critique des étudiants inscrits dans un programme de premier cycle en orthophonie

KEYWORDS
CRITICAL THINKING
TEACHING METHODS
COMMUNICATION SCIENCES AND DISORDERS

Richard J. Morris
Alexandra E. Brockner
Sarah E. Coleman

Abstract

The purpose of this study was to evaluate the effectiveness of two methods for teaching critical thinking skills to communication sciences and disorders students. It was hypothesized that a short course of critical thinking training would result in improved student scores on critical thinking assessments, and that students taught using a mixed instruction method would exhibit greater improvement in their critical thinking skills. Pre- and post-test critical thinking assessments were compared for students who completed 10 weeks of critical thinking instruction. The students were instructed using either (a) a mix of direct instruction on critical thinking concepts along with problem-based learning communication sciences and disorders examples or (b) infused problem-based learning critical thinking instruction with communication sciences and disorders based problems. The pretests and posttests consisted of a general and content specific critical thinking assessment. All of the students exhibited improved scores on both critical thinking measures. In addition, the students who received the mixed instruction method exhibited greater improvements. The greatest improvements for all students occurred for the trained critical thinking skills. These results indicate that both mixed and infused instruction can be effective in teaching students critical thinking skills; however, the mixed instruction was more effective.

Richard J. Morris
Florida State University, Tallahassee, FL, USA
Alexandra E. Brockner
Northwestern University, Evanston, IL, USA
Sarah E. Coleman
Florida State University, Tallahassee, FL, USA

Abrégé

L'objectif de cette étude était d'évaluer l'efficacité de deux méthodes d'enseignement pour développer l'esprit critique des étudiants inscrits dans un programme d'orthophonie. Les auteurs ont avancé l'hypothèse que l'ajout d'un bref cours sur des concepts d'esprit critique permettrait d'améliorer les scores obtenus par les étudiants à des évaluations de leur esprit critique. Ceux-ci ont également suggéré qu'une plus grande amélioration de l'esprit critique serait observée chez les étudiants inscrits dans un cours utilisant une méthode d'enseignement mixte. Les scores obtenus aux évaluations réalisées avant et après les 10 semaines de cours pour évaluer l'esprit critique des étudiants ont été comparés. Les cours donnés aux étudiants utilisaient soit (a) une méthode d'enseignement mixte, incluant un enseignement direct des concepts d'esprit critique et des exemples d'apprentissage par problèmes spécifiques à l'orthophonie, ou (b) une méthode où l'esprit critique était enseigné dans un contexte d'apprentissage par problèmes spécifiques à l'orthophonie. Les évaluations réalisées avant et après les 10 semaines de cours consistaient en une évaluation générale et spécifique à l'orthophonie de l'esprit critique. Les scores de tous les étudiants aux évaluations de leur esprit critique se sont améliorés. Il faut néanmoins noter que les scores des étudiants ayant reçu un enseignement mixte se sont davantage améliorés. Ajoutons également que les scores des concepts d'esprit critique ciblés dans la présente étude se sont davantage améliorés, et ce, pour tous les étudiants. Ces résultats indiquent que les deux méthodes d'enseignement investiguées dans la présente étude peuvent être efficaces pour développer l'esprit critique des étudiants. Or, la méthode d'enseignement mixte semble plus efficace.

An emphasis on evidence-based practice in the teaching and practice of communication sciences and disorders (CSD) in recent years has brought increased focus on the critical thinking (CT) skills of CSD students and professionals. For example, Finn, Brundage, and DiLollo (2016) stated that CT knowledge and skills provide a framework for quality decision making and can be considered a core competency for implementing interprofessional practice. These authors also reported that the American Speech-Language-Hearing Association recognizes that CT knowledge and skills are essential for effective clinical education. Similarly, Gunter and LeJeune (2015) explained that CSD clinicians need CT knowledge and skills in order to develop and maintain ethical practices such as the commitment to maintain and enhance professional competence, accurate representation of information, and accountability for professional standards. Essential attributes of CT, such as skepticisms and insistence on evidence to support statements, can help communication disorder professionals assess and implement the most effective treatment strategies (Finn, Bothe, & Bramlett, 2005).

To illustrate the need for CT skills in CSD clinical practice, Kamhi (2011) compared researchers in the scientific community with CSD clinicians. He contended that researchers are aware of the fallibility of scientific knowledge and the role of the scientific community in determining the reliability, validity, and importance of research findings. In contrast, he stated that most clinicians operate individually when diagnosing and treating clients and are seldom required to justify their clinical decisions to their peers. Kamhi indicated that the lack of inquiry within the clinician community lends itself to overconfidence in one's own ability and practices and a lack of recognition for the fallibility of such practices. Thus, clinicians need to be trained to question their clinical practices and to skeptically evaluate new practices. They should develop a consistent, hierarchical, data-driven approach to clinical decision making (Kamhi, 1984). Kamhi (2011) suggested that clinicians use Dollaghan's (2007) version of evidence-based practice by incorporating practice-based evidence as a means for skepticism toward their own practices and new ones. Such assessment of evidence and subsequent problem solving is a form of CT.

Finn and colleagues (Finn, 2011; Finn et al., 2016) discussed the importance of CT for the development of clinical skills in CSD students. They expressed concern that an emphasis on training CSD students to use evidence-based practices is necessary but not sufficient for these student clinicians to avoid confirmation bias and other thinking errors that can affect clinical decision making. They stated that interpretation, evaluation, and metacognition are CT skills that CSD students need in order to engage in more effective thinking about clinical practices (Finn, 2011; Finn et al., 2016). These skills are similar to the abilities reported as essential to CT in other disciplines, including analysis, evaluation, self-regulation, the ability to distinguish relevant from irrelevant information, and the ability to pose questions whose answers will help to broaden and focus understanding of an issue (Uba, 2008; Yang & Chou, 2008). This purposeful analysis requires skepticism, self-discipline, and awareness of thinking errors (Abrami et al., 2008; Finn et al., 2016; Gunter & LeJeune, 2015). Finn (2011) concluded that CT needs to be directly taught to CSD students as they are unlikely to develop the necessary thinking skills indirectly.

Critical Thinking Instruction

Student and clinician success within and beyond the classroom depends on the teaching and development of CT skills and dispositions (Semerci, 2005; Uba, 2008; Yang & Chou, 2008). However, different concepts of CT instruction result in varying curricular designs and educational approaches within and across disciplines (Thomas & Lok, 2015). Those who consider CT skills to be generic abilities that apply across different content areas state that these skills and dispositions can be taught in stand-alone courses without concern as to the content used to develop them (Royalty, 1995; Sá, West, & Stanovich, 1999). Whereas others contend that these skills are subject or content dependent and that CT skills are best learned as a component of courses centred on the students' academic interests (Halliday, 2000; Smith, 2002). A meta-analysis revealed greater effectiveness for teaching CT in a content dependent context (Abrami et al., 2015).

Abrami and colleagues (Abrami et al., 2015; Abrami et al., 2008) found that those who hold a generic skills perspective of CT tend to support explicitly teaching the underlying skills and dispositions, while those who hold a content based perspective tend to support embedding the CT skills into course content and providing implicit instruction of CT skills. From these two educational approaches come four instruction techniques: general, infused, immersed, and mixed (Abrami et al., 2015; Abrami et al., 2008; Ennis, 1989). Abrami et al. (2008) and Abrami et al. (2015) described these techniques as follows. The *general technique* involves teaching CT abilities separately from any other subject matter. When using the *infused technique*, the instructor uses familiar subject matter as the foundation for teaching CT in the context of the material and CT goals are explicitly taught. The *immersion technique* includes the same teaching structure as the

infused method except that the CT goals are not explicitly taught. Finally, when using the *mixed technique*, the instructor combines the general technique with either the immersion or infused technique. Although these instruction techniques have been evaluated for other clinical fields (Choi, Lindquist, & Song, 2014; Coker, 2010; Macpherson & Owen, 2010; Oja, 2010; Prosser & Sze, 2014), assessment of the effectiveness of CT instruction techniques with CSD students is needed.

An understanding of the teaching techniques can help instructors determine how they might teach CT, but they need to appreciate some of the challenges in this instruction. For example, Thomas and Lok (2015) said that CT skills and knowledge acquisition are necessary but not sufficient for students to use evaluative reasoning and metacognition; the disposition to utilize CT knowledge and skills consistently is required. Developing these thinking skills and dispositions can appear to be a daunting task for instructors. Therefore, instructors should be aware that defining, assessing, and teaching CT skills and dispositions should be undertaken with the understanding that developing these skills and dispositions will need to be program goals over multiple courses (Wendland, Robinson, & Williams, 2015). Wendland et al. (2015) said that when students are developing and utilizing CT skills and strategies they need multiple opportunities to question the information and skills they are taught as well as encouragement to find alternative perspectives. CSD students need to recognize that a skeptical, inquisitive approach to knowledge and clinical situations will help them make better clinical decisions (Apel, 2011; Finn, 2011; Kamhi, 2011; Orlikoff, Schiavetti, & Metz, 2015).

Pedagogical Methods for Teaching Critical Thinking

Suggestions for teaching CT skills and dispositions within an embedded educational approach include pedagogical methods such as problem-based learning, team-based learning, case presentations, and a variety of mapping activities (Day & Williams, 2000; Dochy, Segers, Van den Bossche, & Gijbels, 2003; Johnstone & Otis, 2006; Leahy, Dodd, Walsh, & Murphy, 2006; Mok, Whitehill, & Dodd, 2008; Tiwari et al., 2006; West, Pomeroy, Park, Gerstenberger, & Sandoval, 2000). Meta-analyses comparing health care student outcomes from problem-based learning and traditional classrooms indicate that problem-based learning is more effective than didactic presentations in the development of psycho-motor, affective, and cognitive skills as well as better learning of clinical skills (Prosser & Sze, 2014; Shin & Kim, 2013). Although the authors of several studies have suggested the aforementioned pedagogical methods to help students develop CT, few data exist that indicate student thinking changes as a result of these methods. Thus, a need exists to empirically evaluate the effectiveness of these methods.

It is possible that problem-based learning is a more effective pedagogical method for instructing advanced students. For example, nursing and CSD student learning styles appear to develop in a manner consistent with CT skills (Elliott & Hennessey, 2001; Shin & Kim, 2013). These changes result in greater development of CT skills among graduate students after completing courses using problem-based learning than in undergraduate students who completed similar courses (Shin & Kim, 2013). Thus, a foundation of both content specific knowledge and CT skills may be needed for problem-based learning activities to be most effective, and the pedagogical method used for CT instruction may need to evolve as the students mature.

Gaps in the Current Literature

Several authors have indicated the need for CSD students to develop their CT knowledge and skills in order to effectively select and implement evidence-based practices and treatment techniques (e.g., Finn, 2011; Finn et al., 2005; Finn et al., 2016; Gunter & LeJeune, 2015; Kamhi, 2011; Orlikoff et al., 2015). Other authors have proposed pedagogical methods designed to provide students the opportunity to develop their CT knowledge and skills (e.g., Day & Williams, 2000; Dochy et al., 2003; Johnstone & Otis, 2006; Leahy et al., 2006; Mok et al., 2008; Tiwari et al., 2006; West et al., 2000). However, a need exists for evidence on the effectiveness of these pedagogical methods for the development of CT skills in CSD students. Before CSD faculty members adjust their teaching style and curriculum, they need evidence that these pedagogical methods are effective.

Objectives

As previously stated, instructors across academia and in CSD have increased interest in teaching CT skills and dispositions. However, few data are available concerning the effectiveness of teaching CT skills in CSD programs. In addition, opinions differ on the best educational approach, instructional technique, and pedagogical method for teaching these skills.

Data on the effectiveness of teaching CT skills in CSD programs are needed. In addition, a comparison of classroom instruction techniques may help guide CSD instructors to effectively teach CT skills. Thus, this study had two aims. The first aim was to determine the effectiveness of a short CT course in improving the CT skills of CSD undergraduate students. It was hypothesized that

CT instruction would positively affect the thinking skills of the CSD students. The second aim was to compare mixed CT instruction with infused CT instruction in a group of undergraduate CSD students. Both of these instruction techniques use the content dependent approach reported to be more effective for teaching CT (Abrami et al., 2015). Based on the findings of evolving thinking skills among students (Elliott & Hennessey, 2001; Shin & Kim, 2013) and the reported need for foundation CT knowledge to develop applied CT skills and dispositions (Davies, 2013), it was hypothesized that the undergraduate students taught via the mixed instruction method would exhibit greater improvement in CT skills.

Method

The Florida State University Human Subjects Committee approved the study design and the consent form on August 24, 2015, with approval number HSC # 2015.15827. The study was completed using the approved design. The study was a one-shot between groups pretest-posttest comparison.

Participants

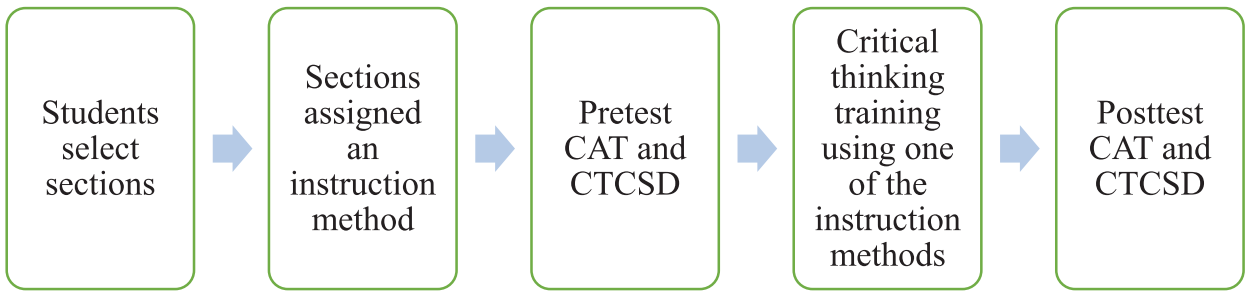
Sixty-seven undergraduate students (1 man, 66 women) aged 18–22 years who were enrolled in a mandatory CSD course served as the participants for this study. The students who participated were 67 of the 85 who had applied and been admitted to a limited access upper division CSD program that required at least a 3.3 grade point average for admission. The other 18 students opted to not participate in the study. The participating students were enrolled in four laboratory sections of a single course. The students enrolled in the sections without knowledge of the study. No attempt was made to control which students enrolled in any of the sections. Therefore, they assigned themselves to the specific sections based on their own personal criteria and the timing of when the sections reached maximum enrollment. This course is taught during the fall semester of the junior year, so the students were in their first semester of CSD undergraduate course work. Only the 85 students enrolled in the course were eligible to participate in the study. At the beginning of the semester each participant signed the approved informed consent form.

Procedures

Figure 1 shows the sequence of procedures to complete the study. As can be seen, the students selected a course section, the sections were assigned an instruction method, the pretesting was completed, the CT instruction occurred, and then the posttesting was completed.

Instruction techniques and materials. Groups were created by designating two of the sections to engage in a mixed direct and infused instruction technique and two of the sections to engage solely in an infused instruction technique. A problem-based learning pedagogical method was the infused instruction technique used for both groups. The determination of the instruction approach to use in each course section was made without any knowledge of the enrolled students. Demographic data on the students in the two instruction groups are shown in **Table 1**. In order to determine if the participants in the two instruction groups were academically equivalent, their grade point averages were compared using a *t* test. The *t* test indicated that the two groups did not differ significantly, $t(31) = 0.15$, $p > .05$. Thus, for the purpose of this study, the students in the two instruction groups were considered to be academically equivalent.

Figure 1



Flow diagram of procedures. The pre- and post-tests were the Critical-thinking Assessment Test (CAT) and Critical Thinking in Communication Sciences and Disorders (CTCSD).

Table 1

Demographic Data of Student Participants for Age, Gender, and Grade Point Average

Instruction Group	Age in Years	Gender	GPA *M* (*SD*)	GPA Range
Problem-based learning	18–21	31 women, 1 man	3.73 (0.15)	3.44–4.00
Mixed instruction	19–22	35 women	3.79 (0.18)	3.53–4.00

Note. GPA = Grade Point Average.

A designated instructor, author AB, taught the mixed instruction sections. Another instructor, author SC, taught the infused instruction sections. Both instructors were trained by the first author on the pedagogical techniques to use. The CT instruction occurred during ten 50-minute sessions—an introductory session and nine training sessions. The material and clinical cases used during the instruction were unique from those in the tests used to assess the students' CT knowledge and skills.

The topics of CT explicitly addressed during the instruction for this study included logical fallacies and thinking errors, problem solving, and evaluating causal claims. According to Facione (2015), these instructional topics fall into three of six categorical skills of critical thinking: evaluation, inferencing, and self-regulation. Examples of evaluation include analyzing the credibility of claims and the facts which support them (Facione, 2015). Examples of inferencing, according to Facione, include drawing conclusions from given information, ruling out conclusions from given information, and considering alternatives. The examples provided for self-regulation include assessing one's own methodology before committing to an answer (Facione, 2015).

The mixed instruction included presentations by the instructor and small group discussions on each of the topics during three 50-minute class sessions. Student evaluations included a set of short answer responses concerning the thinking skills and two concept maps. The concept maps depicted the student thinking on two case studies, one focused on problem solving and the other on decision making skills.

For example, the training on effective problem solving began with instruction on creating a concept map with examples and simple practice problems. Then author AB presented a clinical case issue about the Individualized Education Program of a school-aged child. The students were directed to individually write their solution to the problems in the case; then they were directed to write the thinking procedures they used to solve the problems. After that, they formed groups of two or three students and compared the strategies they used. Next, they had a full class discussion of the strengths and weaknesses of the various strategies. Then author AB presented an organization for thinking about problem solving based on the writing of Beyer (1987). The students returned to their small groups and compared their strategies to Beyer's. Then a second case was presented that involved an older person with a hearing loss. Again they wrote out how they would solve the problem followed by writing out the thinking process used in developing the solution. They discussed the second case in their small groups. Then author AB answered questions regarding the problem-solving strategies and how to develop a concept map of their problem solving for this case. Then a third case study about a post-cardiac surgery patient with a swallowing problem was presented and the students wrote out solutions to the presented problem. Each student then created a concept map of her problem-solving strategy that was submitted for a grade.

The infused instruction included three clinical cases created by the first author. The first clinical case included thinking errors by a parent and clinician for the students to explore. The students were encouraged to develop a problem-solving strategy for structuring an evaluation of the communication problem in the second clinical case. Finally, the third clinical case included a variety of professionals discussing the cause of a communication problem. The discussions of the cases were structured with times for dyad, small group, and large group discussion. The instructor was trained to reflectively respond to the students and to minimize her input to the students' developing CT skills and dispositions. At the end of each section, the groups of students submitted concept maps to represent how they

conceptualized the situation. In addition, the students were encouraged to ask questions and to complete independent research to understand each clinical case.

For example, the second clinical case involved a young woman who was having difficulties singing. The case background included information about vocal demands in her work environment, her singing, and her personal life. During the first session the students read the clinical case and wrote a list of questions/issues. They were then directed to discuss their questions/issues with two or three other students in the section. The groups of students then provided questions for author SC to answer for all of the students in the section. As noted above, author SC would respond in a manner to help them focus on the problem-solving strategies they used, such as, "What was the focus of your thinking when you developed that question? Might there be another way to think about the material that could lead you to a different question?" or "Since that point might not be relevant to solving the issue, how might you approach the case to develop more relevant ideas?" In the next session, the focus tended to be on the research students had done on the topic to help them determine appropriate and relevant problem-solving strategies. In the third session the students brought their individual concept maps of how they structured the known information, what they still needed to know, and what evaluation tools and methods they would use. They shared their concept maps in small groups and discussed their similarities and differences. After author SC answered the students' questions and discussed problem solving with them, the students recreated individual concept maps and submitted them.

Critical thinking assessments. Baseline measurements of the students' CT skills were taken during the first week of class. These same assessments were repeated at the beginning of the following semester to collect post-treatment data. A 60-minute period was allotted for completing each of the CT assessments. The students finished the assessments in 45–60 minutes.

Measurements were taken using a general CT assessment, the Critical-thinking Assessment Test (CAT; Center for Assessment and Improvement of Learning, n.d.), and a content specific CT assessment, Critical Thinking in Communication Sciences and Disorders (CTCSD; Morris, Gorham-Rowan, Coston, & Scholz, 2014). The CAT contains 15 items, 14 of the items are prompts for short essay responses and one is a prompt for mathematical calculation. The 15 items assess four CT skills: evaluating and interpreting information (8 items), problem solving (8 items), creative thinking (6 items), and effective communication (9 items). Four of the items assess evaluating and interpreting information only, one item assesses problem solving only, five of the items assess two of the CT skills, and the remaining six assess three of the CT skills. Stein and Haynes (2011) reported that performance on the CAT was significantly correlated with performance on other tests of critical thinking, $r = .65$ with the California Critical Thinking Skills Test (Facione, 1990) and $r = .69$ with the Critical Think Module of the Collegiate Assessment of Academic Proficiency (ACT Inc., 2000). These findings demonstrate the content validity of the CAT. Higher education faculty members from a broad range of disciplines who score the CAT have agreed that the items on the CAT assess CT skills with a range of 80% to 100% agreement for each of the items, indicating face validity of the assessment (Stein & Haynes, 2011). These trained faculty members have a high inter-judge reliability of $r = .92$ (Stein, Haynes, & Redding, 2007). The items on the CAT exhibited high internal consistency with a Cronbach alpha of $\alpha = 0.82$ (Stein et al., 2007).

The CTCSD consists of 14 prompts for short answer responses and two prompts for mathematical calculations. **Table 2** displays the target CT skills for the items in the CAT and CTCSD. Student performance on the CTCSD has been highly correlated with performance on the CAT with $r = .793$ ($p < .01$) and on the Ennis-Weir Critical Thinking Essay Test (Ennis & Weir, 1985) with $r = .629$ ($p < .01$) indicating the content validity of the CTCSD (Morris et al., 2014). The scorers of the CTCSD had a high inter-judge reliability of $r = .95$.

The two tests have similar prompt styles and scoring systems. An example of a prompt from the CTCSD reads:

> In the late 1990s a new modality of treatment for oropharyngeal dysphagia was approved by the FDA. This treatment involves electrical stimulation of the neck muscles via surface electrodes (NMES). Speech language pathologists (SLPs) working in acute care hospitals and rehabilitation facilities have observed that 85% of their patients who received NMES as part of their post-stroke treatment exhibited improved swallow. These SLPs say that NMES is an effective tool for improving swallow function among post-stroke people with dysphagia.
>
> *Provide two alternative explanations that might explain the improvements in swallow behavior among these patients.*

These prompts were scored on the quality of the explanations and reason underlying the explanations. The scoring on the individual prompts ranged from a two-point 0–1 range for yes/no questions and mathematical

Table 2

Skills Targeted by Each Critical-Thinking Assessment Test and Critical Thinking in Communication Sciences and Disorders Assessment Item

CAT Item	CTCSD Item	Target Skill
1	10	Summarize the pattern of results in a graph without making inappropriate inferences
2	1	Evaluate how strongly correlational-type data supports a hypothesis
3	2, 12	Provide alternative explanations for a pattern of results that has many possible causes
4, 7	4	Identify additional information needed to evaluate a hypothesis
5	8	Evaluate whether spurious information strongly supports a hypothesis
6	9	Provide alternative explanations for spurious associations
8	11	Determine whether an invited inference is supported by specific information
9	3, 7	Provide relevant alternative interpretations for a specific set of results
10	14	Separate relevant from irrelevant information when solving a real-world problem
11	13	Use and apply relevant information to evaluate a problem
12	6a, 6b	Use basic mathematical skills to help solve a real-world problem
13	5	Identify suitable solutions for a real-world problem using relevant information
14	9	Identify and explain the best solution for a real-world problem using relevant information
15	15	Explain how changes in a real-world problem situation might affect the solution

Note. CAT = Critical-thinking Assessment Test; CTCSD = Critical Thinking in Communication Sciences and Disorders.

calculations to three- and four-point 0–2 and 0–3 ranges for questions like the one above that had a range of 0–3.

Analysis

Once the CT pre- and post-tests were completed, participants' responses were scored for both assessment tools. The people scoring the two assessments underwent training to develop inter- and intra-judge reliability. For a score to be counted, two scorers had to agree on the points awarded for the written response. If agreement was not reached between the first two scorers, the third scorer read and scored the test item, with the students' scores always requiring that two scorers agree on the score for every item. The faculty members who scored the CAT were trained by instructors from the Center for Assessment and Improvement of Learning, who also rescored the test for reliability. The faculty members completing the scoring only knew that the assessments were completed by CSD students and were not aware of this study. For the CTCSD, the first author trained the other two authors until they could consistently score items and report their criteria for the scoring of the responses. When necessary, the first author also served as the third scorer.

The numbering system of the student responses provided by the Center for Assessment and Improvement of Learning for the university scoring of the CAT did not provide for separating the sections of the course. Thus, the CAT scores provided information on the CT skills exhibited pre- and post-training, but separate scores were not available for the students who participated in the mixed or infused instruction sections. The CTCSD responses were identified by a participant number only so that the scorers would not know whose responses they were reading. After scores were assigned to all of the completed CTCSDs, the scores were separated between the mixed and infused instruction sections based on a digit in the participant numbers.

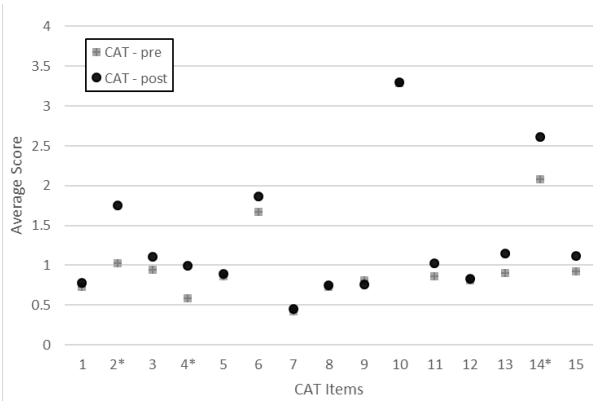
To determine any changes in the students' scores Pillai's Trace MANOVA was used. Pillai's Trace was selected because it is regarded as the most powerful and robust of the four MANOVA test statistics (Pillai, 2004). This statistic was completed as part of the SPSS repeated measure ANOVA routine (IBM Corp, 2015). The between-subjects effect of the repeated measures ANOVA was used to determine any instruction group effect.

Results

The results of the student performance on the CAT provided by the Center for Assessment and Improvement of Learning indicated that the students in both instruction groups exhibited significant overall CT skill improvement on the posttest (see **Figure 2**). The mean total score improved from 16.55 to 19.28, $p < .001$, with an effect size of .62. In addition, the Center for Assessment and Improvement of Learning reported improvement on three of the 15 assessment items (i.e., 2, 4, and 14), two of which relate to problem solving and one to evaluating and interpreting information. The average of the student scores improved from 0.59 to 0.99, $p < .01$, with an effect size of .38 for item 4, and improved from 2.08 to 2.61, $p < .05$, with an effect size of .29 for item 14. For item 2, the average of the students' scores improved from 1.03 to 1.75, $p < .001$, with an effect size of .72. The average student scores for 11 of the other 12 CAT items improved, but not significantly (see **Figure 2**).

Similarly, the students in both instruction groups exhibited an overall improvement on the posttest of the CTCSD in comparison to the pretest. As shown in **Figure 3**, the average posttest scores for all but one of the CTCSD prompts (i.e., prompt 8) was higher than the matched pretest score. In addition, the average score of the students in the infused instruction group did not change between the pretest and posttest for one prompt

Figure 2

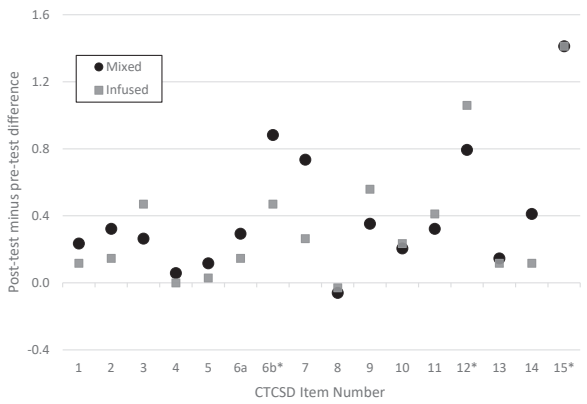


Average Critical-thinking Assessment Test scores from assessments of participants before and after the training sessions. Item with * significantly improved ($p < .05$). CAT = Critical-thinking Assessment Test.

(i.e., prompt 4). The average posttest scores on the CTCSD were significantly higher than the average pretest scores as indicated by the Pillai's Trace MANOVA, $F(1, 63) = 199.73$, $p < .01$, $\eta_p^2 = .760$. **Figure 3** shows similar results for the CAT, the student scores varied among the CTCSD questions, as shown by the Pillai's Trace MANOVA comparison across the assessment items, $F(15, 49) = 86.02$, $p < .01$, $\eta_p^2 = .963$. The students exhibited the greatest improvements on the CTCSD prompts that were associated with the content of the instruction. This finding indicates an association between participant responses to certain prompts on the CTCSD and the average score differences between the pre- and post-tests, $F(15, 49) = 11.96$, $p < .01$, $\eta_p^2 = .785$. CTCSD item 6b evaluated the students' mathematical skills. The mean on this item improved from 0.83 to 1.71 among the students in the mixed instruction group and from 0.87 to 1.39 among the students in the infused instruction group with an overall $p < .001$. Item 12 evaluated their ability to provide alternative explanations for a pattern of results. The mean on this item improved from 1.59 to 2.38 among the students in the mixed instruction group and from 0.84 to 1.97 among the students in the infused instruction group with an overall $p = .002$. Finally, item 15 evaluated their ability to explain how changes in a problem might affect the solution. The mean on this item improved from 1.11 to 2.53 among the students in the mixed instruction group and from 0.35 to 1.81 among the students in the infused instruction group with an overall $p < .001$.

Although the specific focus of the items with significant improvement on the CAT and CTCSD differed, the items assessed the students' ability to evaluate and interpret information. The exception was the improvement in mathematical skill on the CTCSD. In general, the greatest student improvements occurred for similar CT skills on both assessments.

Figure 3



Differences in average Critical Thinking in Communication Sciences and Disorders scores between the assessments of the participants before and after the training sessions for the students who had the mixed instruction and those who had the problem-based learning instruction. Item with * significantly improved ($p < .05$). CTCSD = Critical Thinking in Communication Sciences and Disorders.

Although the mean scores for both instruction groups improved on the CTCSD posttest, a between-group ANOVA comparison of the two instruction groups revealed that the average of the students' scores in the mixed instruction group exhibited more improvements than those of the students in the infused instruction group, $F(1, 63) = 13.13$, $p < .01$, $\eta_p^2 = .172$. The average pretest to posttest change for the two teaching styles can be seen in **Figure 3.** The students in the mixed instruction group exhibited greater average pretest to posttest improvement for nine of the CTCSD prompts (i.e., 1, 2, 4, 5, 6a, 6b, 7, 13, and 14). The change was the same for one prompt (i.e., prompt 15), and the students in the infused instruction group exhibited greater average improvements for six of the prompts (i.e., 3, 8, 9, 10, 11, and 12). In addition to exhibiting greater improvement for more of the prompts, the students' average scores in the mixed instruction group also showed greater improvement for their individual CTCSD prompt scores ($M = 0.41$) than the students in the infused instruction group ($M = 0.35$).

The effect sizes for both the main effects as well as the interaction effect were in the high range (Cohen, 1988). This finding indicates that all of the statistical effects explain a high proportion of the observed variability in the data.

Discussion

Effectiveness of Critical Thinking Training

The findings from this study indicate that a 10-week course utilizing either a mixed or infused instruction technique can be effective in improving CT skills in CSD students. These results are consistent with the findings from other higher education disciplines of statistically significant CT improvements after instruction using a problem-based learning pedagogical method (Butchart et al., 2009; Casotti, Rieser-Danner, & Knabb, 2008; Lombard, 2008; Reynolds & Hancock, 2010).

The responses by the students in both instruction groups indicated a modest improvement in selected CT skills from the instruction, which corresponds with previous reports (e.g., Abrami et al., 2015). These results were found in the outcomes of both the general knowledge CAT and the content specific CTCSD. The students exhibited improvements for the assessment items that evaluated the CT skills that were trained during the semester. In contrast, their posttest scores for the other items on both tests improved slightly, but were similar to the pretest scores. Since the training in both groups targeted similar CT skills, the other items could be regarded as control items. Improvements in the targeted skill items may indicate the effectiveness of the training methods in teaching CT skills to these students. In addition, the contrast between the trained and untrained assessment items reveals that the higher scores were more content specific than general learning or experience would explain. However, the possibility exists that the improvements reflected the acquisition of skills for responding to specific prompts. Thus, the improvements may not indicate a change in generic thinking skills, but specific content knowledge.

As noted previously, the topics of CT explicitly addressed during the instruction for this study included logical fallacies and thinking errors, problem solving, and evaluating causal claims. These topics are included in three of the six categorical CT skills that Facione (2015) listed: evaluation, inferencing, and self-regulation.

These CT skills align with the target skills associated with the CTCSD prompts with the greatest posttest improvement—6b: use basic mathematical skills to help solve a real-world problem (self-regulation), 7: provide relevant alternative interpretations for a specific set of results (inferencing), 12: provide alternative explanations for a pattern of results that has many possible causes (inferencing, evaluation), and 15: explain how changes in a real-world problem situation might affect the solution (inferencing). These findings suggest that the CT instruction was associated with improvements in these students' CT assessment scores. This finding concurs with previous research on the effectiveness of teaching CT skills (Abrami et al., 2015; Abrami et al., 2008; Glaser, 1941).

The exception to this pattern was the improvement in mathematical skills exhibited by the students on the CTCSD posttest. This difference may have been a result of increased student comfort with the assessment and reduced test related anxiety.

Effectiveness of Instruction Technique

The results of the instruction technique comparison in this study agree with the previous reports (Abrami et al., 2015; Abrami et al., 2008) that the mixed method of instruction is an effective method for initial teaching of CT skills. These findings are similar to those of Stoiber (1991) who found that direct instruction in the use of reflective thinking for solving discipline specific problems resulted in improved evaluation of information and problem-solving skills. These results also indicate that the explicit CT instruction in this study appears to be more effective in teaching these undergraduate CSD students CT skills than the implicit instruction method used with CSD students (Grillo, Koenig, Gunter, & Kim, 2015). These authors reported limited improvements in CT skills after problem-based learning instruction. In contrast to Grillo et al.'s (2015) findings, results from other clinical fields also indicate that short-term problem-based learning instruction is associated with improved CT skills (Choi et al., 2014; Coker, 2010; Macpherson & Owen, 2010; Oja, 2010). More research is needed on the effectiveness of all CT pedagogical methods when teaching CSD students. Future studies can use the examples of undergraduate and graduate CSD courses provided by Finn et al. (2016). These authors also reported a variety of pedagogical methods designed for CSD students to improve CT skills (Finn et al., 2016).

As previously noted, an infused or immersed technique of problem-based learning instruction without direct CT instruction may not be the best method when teaching new content (Butchart et al., 2009; Casotti et al., 2008; Lombard, 2008; Reynolds & Hancock, 2010). Participants in this study, like those in other studies that successfully used problem-based learning (e.g., Butchart et al., 2009), expressed concern about their difficulties in determining what was expected of them and the lack of confidence they felt when completing the concept maps.

Limitations

The participants in this study may have been influenced by outside factors, such as the CSD content in other courses. At the beginning of the study they had no formal CSD training and may have only known of CSD treatments through family members or generally available internet/media information. This particular effect may be seen in the improvement of CTCSD item 15 whose target skill was to explain how changes in a real-world problem might affect the solution. Item 15 centred on hearing loss and amplification, and at the time of posttest administration the participants were completing the second week of an Introduction to Audiology course. The content that the students learned in the other course may have informed their answers more than the CT instruction. A time-series type of study with a sequence of CT skills taught over a series of semesters could be a way to more thoroughly evaluate the effectiveness of teaching CT skills.

Another possible issue with this study could be the competence of the instructors. The improvements on the prompts that addressed inferencing skills with little change for the prompts that addressed interpretation and analysis could imply that the instruction was varied in quality as a result of the instructor's own CT strengths or weaknesses. Although the instructors were trained on the methodology used for this study, no assessment was given to gauge their skills. Previous work indicates that instructor quality affects CT instruction effectiveness (Abrami et al., 2008).

A third limitation was that two of the study's authors (i.e., AB and SC) were both instructors and scorers of the CTCSD. Although the CTCSD responses were identified only by a participant number, AB and SC knew the numbering scheme and could identify the participant's section. However, as reported above, AB and SC exhibited high inter-judge reliability. Thus, they did not exhibit a tendency to score the participants from their sections higher than they scored the other participants. In the future, it will be better to have scorers who have no other involvement with the study and do not know the participants' training group.

Finally, since the CAT and CTCSD have a similar structure they may assess the same aspects of CT skills. Pairing these

two assessments with an assessment that has a different structure, such as the California Critical Thinking Skills Test, might improve the validity of the findings. In addition, the lack of CAT data that could separate the CT instruction methods limits the information on differences in student outcomes. Future studies should provide a method for separating the data. Another suggestion is that future studies involving CT assessment of CSD students should include a qualitative analysis of student opinions concerning the CT training and assessment.

Critical Thinking and Communication Sciences and Disorders Training

Further research is required to demonstrate the effectiveness of CT instruction in CSD programs and to more clearly define the relationship between specific methods of CT instruction and improved CT skills among CSD students. Such studies could include a longer course of training and use of other pedagogical techniques.

Critical thinking skills are vital to speech-language pathologists and audiologists as they provide a quality thinking structure to assist in the decision making and problem solving involved in the evaluation and treatment of clients (Finn et al., 2005; Gunter & LeJeune, 2015). As Orlikoff et al. (2015) stated, CT is a fundamental aspect of clinical practice in communication disorders. By working to improve these thinking skills and dispositions through targeted classroom activities, students can hone their ability to reevaluate their thought processes and relevant information in order to solve a clinical problem. By doing so, they can be better prepared to make accurate diagnoses and create appropriate treatment plans. The improvement of these skills should help these students become clinicians who will recognize the need to be current in their understanding of communication disorders as well as the evaluation and treatment of the disorders. With these thinking skills and attributes they should be willing and able to work and re-work complex clinical problems until they find the most functional solutions for their clients.

The pedagogical implications from the current and previous studies indicate a sequence of instruction to help students develop their CT skills and dispositions (Bailin & Battersby, 2015; Byrnes & Dunbar, 2014; Shin & Kim, 2013; Wendland et al., 2015). These results indicate that a mixed instruction method may be the better method for an initial course in which students directly learn CT skills. Future research may reveal that the problem-based learning pedagogical approach may be more effective in a subsequent course to help the students develop their CT skills into CT dispositions. In addition to the sequencing of CT courses for better student learning, instructors should be aware that students often have difficulty grasping the purpose of problem-based learning activities. Therefore, the instructor needs to invest time explaining how problem-based learning works in order to increase the effectiveness of the course (Prosser & Sze, 2014). In addition, student response to the pedagogical techniques needs to be known. Future studies should ask the students for their opinions of the techniques and determine (a) what they felt they learned, (b) if they found benefit to the material, and (c) if they found benefit to the pedagogical techniques.

In conclusion, the pursuit to advance CT instruction for CSD students has achieved significant notice but needs wider implementation. The present and past studies indicate teaching strategies and techniques associated with improved CT assessment scores. As the need for these skills have been established, routine implementation of CT instruction in CSD programs is the logical next step (Finn, 2011). The current study provides evaluation of two teaching strategies for implementing Finn's suggestion. Further studies of CT teaching strategies in CSD courses should provide improved understanding of the best methods to improve thinking strategies among CSD students. Improvements in thinking strategies can be a tool for the increased scientific and skeptical thinking that Kamhi (2011) suggested for improved clinical effectiveness.

References

Abrami, P. C., Bernard, R. M., Borokhovski, E., Waddington, D. I., Wade, A., & Persson, T. (2015). Strategies for teaching students to think critically: A meta-analysis. *Review of Educational Research, 85,* 275–314. doi:10.3102/0034654314551063

Abrami, P. C., Bernard, R. M., Borokhovski, E., Wade, A., Surkes, M. A., Tamim, R., & Zhang, D. (2008). Instructional interventions affecting critical thinking skills and dispositions: A stage 1 meta-analysis. *Review of Educational Research, 78,* 1102–1134. doi:10.3102/0034654308326084

ACT Inc. (2000). *The Collegiate Assessment of Academic Proficiency—Critical Thinking Test*. Iowa City, IA: Author.

Apel, K. (2011). Science is an attitude: A response to Kamhi. *Language, Speech, and Hearing Services in Schools, 42,* 65–68. doi:10.1044/0161-1461(2009/09-0036)

Bailin, S., & Battersby, M. (2015). Teaching critical thinking as inquiry. In M. Davies & R. Barnett (Eds.), *The Palgrave handbook of critical thinking in higher education* (pp. 123–138). New York, NY: Palgrave Macmillan.

Beyer, B. K. (1987). *Practical strategies for the teaching of thinking*. Boston, MA: Allyn and Bacon.

Butchart, S., Forster, D., Gold, I., Bigelow, J., Korb, K., Oppy, G., & Serrenti, A. (2009). Improving critical thinking using web based argument mapping exercises with automated feedback. *Australasian Journal of Educational Technology, 25,* 268–291. doi:10.14742/ajet.1154

Byrnes, J. P., & Dunbar, K. N. (2014). The nature and development of critical-analytic thinking. *Educational Psychology Review, 26,* 477–493. doi:10.1007/s10648-014-9284-0

Casotti, G., Rieser-Danner, L., & Knabb, M. T. (2008). Successful implementation of inquiry-based physiology laboratories in undergraduate major and nonmajor courses. *How We Teach, 32,* 286–296. doi:10.1152/advan.00100.2007

Center for Assessment and Improvement of Learning. (n.d.). *The Critical-thinking Assessment Test.* Tennessee Technological University, Cookeville, TN: Author.

Choi, E., Lindquist, R., & Song, Y. (2014). Effect of problem-based learning vs. traditional lecture on Korean nursing students' critical thinking, problem-solving, and self-directed learning. *Nurse Education Today, 34,* 52–56. doi:10.1016/j.nedt.2013.02.012

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Mahwah, NJ: Erlbaum.

Coker, P. (2010). Effects of an experiential learning program on the clinical reasoning and critical thinking skills of occupational therapy students. *Journal of Allied Health, 39,* 280–286.

Davies, M. (2013). Critical thinking and the disciplines reconsidered. *Higher Education Research & Development, 32,* 529–544. doi:10.1080/07294360.2012.697878

Day, R. A., & Williams, B. (2000). Development of critical thinking through problem-based learning: A pilot study. *Journal on Excellence in College Teaching, 11,* 203–226.

Dochy, F., Segers, M., Van den Bossche, P., & Gijbels, D. (2003). Effects of problem-based learning: A meta-analysis. *Learning and Instruction, 13,* 533–568. doi:10.1016/S0959-4752(02)00025-7

Dollaghan, C. A. (2007). *The handbook for evidence-based practice in communication disorders.* Baltimore, MD: Brookes.

Elliott, J., & Hennessey, N. (2001). Scratching the surface: Speech and hearing science students and their approach to learning. In A. Herrmann & M. M. Kulski (Eds.), *Proceedings of the 10th Annual Teaching and Learning Forum, Proceedings of the 10th Annual Teaching and Learning Forum*. Perth, Australia: Curtin University of Technology.

Ennis, R. H. (1989). Critical thinking and subject specificity: Clarification and needed research. *Educational Researcher, 18*(3), 4–10. doi:10.3102/0013189X018003004

Ennis, R. H., & Weir, E. E. (1985). *The Ennis-Weir Critical Thinking Essay Test.* Pacific Grove, CA: Midwest Publications.

Facione, P. (1990). *The California Critical Thinking Skills Test: College level: Interpreting the CCTST, group norms and sub-scores* (Report No. 4). Millbrae, CA: California Academic Press.

Facione, P. A. (2015). *Critical thinking: What it is and why it counts.* Hermosa Beach, CA: Measured Reasons LLC. Retrieved from https://www.scribd.com/document/357190993/facione-2015-critical-thinking-what-it-is-and-why-it-counts-pdf

Finn, P. (2011). Critical thinking: Knowledge and skills for evidence-based practice. *Language, Speech, and Hearing Services in Schools, 42,* 69–72. doi:10.1044/0161-1461(2010/09-0037)

Finn, P., Bothe, A. K., & Bramlett, R. E. (2005). Science and pseudoscience in communication disorders: Criteria and applications. *American Journal of Speech-Language Pathology, 14,* 172–186. doi:10.1044/1058-0360(2005/018)

Finn, P., Brundage, S. B., & DiLollo, A. (2016). Preparing our future helping professionals to become critical thinkers: A tutorial. *Perspectives of the ASHA Special Interest Groups SIG10, 1*, 43–68.

Glaser, E. M. (1941). *An experiment in the development of critical thinking.* New York, NY: Teachers College, Columbia University.

Grillo, E. U., Koenig, M. A., Gunter, C. D., & Kim, S. (2015). Teaching CSD graduate students to think critically, apply evidence, and write professionally. *Communication Disorders Quarterly, 36,* 241–251. doi:10.1177/1527401145490014

Gunter, C. D., & LeJeune, J. B. (2015). *Critical thinking in clinical practice in speech-language pathology.* Dubuque, IA: Kendall Hunt Publishing.

Halliday, J. (2000). Critical thinking and the academic vocational divide. *The Curriculum Journal, 11,* 159–175. doi:10.1080/09585170050045182

IBM Corp. (2015). *IBM SPSS Statistics for Windows, Version 23.0.* Armonk, NY: Author

Johnstone, A. H., & Otis, K. H. (2006). Concept mapping in problem based learning: A cautionary tale. *Chemistry Education Research and Practice, 7,* 84–95. doi:10.1039/B5RP90017D

Kamhi, A. G. (1984). Problem solving in child language disorders: The clinician as clinical scientist. *Language, Speech, and Hearing Services in Schools, 15,* 226–234. doi:10.1044/0161-1461.1504.226

Kamhi, A. G. (2011). Balancing certainty and uncertainty in clinical practice. *Language, Speech, and Hearing Services in Schools, 42,* 59–64. doi:10.1044/0161-1461(2009/09-0034).

Leahy, M. M., Dodd, B. J., Walsh, I. P., & Murphy, K. (2006). Education for practice in the UK and Ireland: Implementing problem-based learning. *Folia Phoniatrica et Logopaedica, 58,* 48–54. doi:10.1159/000088999

Lombard, B. J. J. (2008). Modelling critical thinking through learning-oriented assessment. *South African Journal of Higher Education, 22,* 1029–1043. doi:10.4314/sajhe.v22i5.42923

Macpherson, K., & Owen, C. (2010). Assessment of critical thinking ability in medical students. *Assessment & Evaluation in Higher Education, 35,* 41–54. doi:10.1080/02602930802475471

Mok, C. K. F., Whitehill, T. L., & Dodd, B. J. (2008). Problem-based learning, critical thinking and concept mapping in speech-language pathology education: A review. *International Journal of Speech-Language Pathology, 10,* 438–448. doi:10.3109/02699206.2013.816781

Morris, R. J., Gorham-Rowan, M. M., Coston, J., & Scholz, K. (2014, Nov.). *Evaluating and teaching causal reasoning skills to determine critical thinking among CSD students.* Paper presented at the annual meeting of the American Speech-Language-Hearing Association, Orlando, FL.

Oja, K. J. (2010). Using problem-based learning in the clinical setting to improve nursing students' critical thinking: An evidence review. *Journal of Nursing Education, 50,* 145–151. doi:10.3928/014834-20201230-10

Orlikoff, R. F., Schiavetti, N. E., & Metz, D. E., (2015). *Evaluating research in communication disorders* (7th ed.). Boston, MA: Pearson Education.

Pillai, K. C. S. (2004). Multivariate analysis of variance. In S. Kotz, C. B. Read, N. Balakrishnan, & B. Vidakovic (Eds.), *Encyclopedia of statistical sciences.* Somerset, NJ: Wiley.

Prosser, M., & Sze, D. (2014). Problem-based learning: Student learning experiences and outcomes. *Clinical Linguistics & Phonetics, 28,* 131–142. doi:10.3109/02699206.2013.820351

Reynolds, J. M., & Hancock, D. R. (2010). Problem-based learning in a higher education environmental biotechnology course. *Innovations in Education and Teaching International, 47,* 175–186. doi:10.1080/14703291003718919

Royalty, J. (1995). The generalizability of critical thinking: Paranormal beliefs versus statistical reasoning. *The Journal of Genetic Psychology, 156,* 477–488. doi:10.1080/00221325.1995.9914838

Sá, W. C., West, R. F., & Stanovich, K. E. (1999). The domain specificity and generality of belief bias: Searching for a generalizable critical thinking skill. *Journal of Educational Psychology, 91,* 497–510. doi:10.1037/0022-0663.91.3.497

Semerci, N. (2005). The effects of problem-based learning on the academic achievement of students in development and learning. *International Journal of Educational Reform, 14,* 415–425.

Shin, I.-S., & Kim, J.-H., (2013). The effect of problem-based learning in nursing education: A meta-analysis. *Advances in Health Science Education, 18,* 1103–1120. doi:10.1007/s10459-012-9436-2

Smith, G. (2002). Are there domain-specific thinking skills? *Journal of Philosophy of Education, 36,* 207–227. doi:10.1111/1467-9752.00270

Stein, B., & Haynes, A. (2011). Engaging faculty in the assessment and improvement of students' critical thinking using the Critical Thinking Assessment Test. *Change: The Magazine of Higher Learning, 43,* 44–49. doi:10.1080/00091383.2011.550254

Stein, B., Haynes, A., & Redding, M. (2007). Project CAT: Assessing critical thinking skills. In D. Deebs & B. Callen (Eds.), *Proceedings of the National STEM Assessment Conference 2006* (pp. 290–299). Washington DC: NSF and Drury University.

Stoiber, K. C. (1991). The effect of technical and reflective preservice instruction on pedagogical reasoning and problem solving. *Journal of Teacher Education, 42,* 131–139. doi:10.1177/0022487198049005009

Thomas, K., & Lok, B. (2015). Teaching critical thinking: An operational framework. In M. Davies & R. Barnett (Eds.), *The Palgrave handbook of critical thinking in higher education* (pp. 93–105). New York, NY: Palgrave MacMillan.

Tiwari, A., Chan, S., Wong, E., Wong, D., Chui, C., Wong, A., & Patil, N. (2006). The effect of problem-based learning on students' approaches to learning in the context of clinical nursing education. *Nurse Education Today, 26,* 430–438. doi:10.1016/j.nedt.2005.12.001

Uba, L. (2008). A deconstructive pedagogy. *Journal of Excellence in College Teaching, 19,* 103–125. doi:10.1080/03004270903206141

Wendland, M. W., Robinson, C., & Williams, P. A. (2015). Thick critical thinking: Toward a new classroom pedagogy. In M. Davies & R. Barnett (Eds.), *The Palgrave handbook of critical thinking in higher education* (pp. 153–168). New York, NY: Palgrave Macmillan.

West, D. C., Pomeroy, J. R., Park, J. K., Gerstenberger, E. A., & Sandoval, J. (2000). Critical thinking in graduate medical education: A role for concept mapping assessment? *Journal of the American Medical Association, 284,* 1105–1110. doi:10.1001/jama.284.9.1105

Yang, Y.-T. C., & Chou, H.-A. (2008). Beyond critical thinking skills: Investigating the relationship between critical thinking skills and dispositions through different online instructional strategies. *British Journal of Educational Technology, 39,* 666–684. doi:10.1111/j.1467-8535.2007.00767.x

Authors' Note

Correspondence concerning this article should be addressed to Richard J. Morris, Communication Science and Disorders, Florida State University, 201 W. Bloxham, Tallahassee, Florida, United States, 32306-1200. Email: rmorris@fsu.edu

Disclosures

No conflicts of interest, financial or otherwise, are declared by the authors.

Busy Toy Designs Reduce the Specificity of Mothers' References to Toy Parts During Toy Play With Their Toddlers

L'utilisation de jouets « chargés » sur le plan visuel réduit la spécificité des références effectuées par les mères sur les parties de ces jouets en situation de jeu avec leur enfant

KEYWORDS
TOY DESIGN
TODDLERS
PARENT–CHILD TALK
REFERENTIAL SPECIFICITY
VOCABULARY
PARENT–CHILD PLAY
TOYS

Daniela K. O'Neill
Taylor J. Deglint
Ashley M. McKinnon
Angela Nyhout
Julianne Scott

Abstract

When a parent is playing with a toy with his or her child, might a toy's "busy" visual design negatively impact the specificity and quality of the parent's talk? In this study, 24 mother–toddler (*M* = 23.5 months) dyads played with both (a) unmodified visually busy commercial toys and (b) modified visually "simple" versions of these commercial toys. Our focus was on the specificity of mothers' 552 references to the main parts of the toys (i.e., the rings of a stacking ring toy and the blocks of a nesting block toy), which was found to be impacted by the toys' visual design. That is, with simple toys, mothers produced a significantly greater proportion of specific references (e.g., *the blue ring*) than non-specific references (e.g., *this/that one*). Indeed, the proportion of specific references was three times greater in play with the simple toys than with the busy toys. Busy toys also reduced the number of references to parts of the toy overall and children's exposure to vocabulary such as colour terms used within specific references. These results underscore that the visual design of toys is an important aspect to consider, particularly in contexts where the goal may be to foster adult–child language and a child's exposure to more information-rich vocabulary terms during toy play with an adult.

Daniela K. O'Neill, Taylor J. Deglint, Ashley M. McKinnon, Angela Nyhout, and Julianne Scott

University of Waterloo, Waterloo, ON, CANADA

Abrégé

Dans une situation de jeu, est-ce que l'utilisation d'un jouet ayant une apparence « chargée » pourrait affecter de façon négative la spécificité et la qualité des échanges entre un parent et son enfant? Dans la présente étude, 24 dyades mère-enfant (*M* = 23,5 mois) ont joué avec (a) des jouets commerciaux non modifiés, dont l'apparence était « chargée », et (b) des versions modifiées et simplifiées sur le plan visuel de ces jouets. Notre attention s'est portée sur la spécificité des 552 références effectuées par les mères sur les parties principales des jouets (c.-à-d. les anneaux d'un jouet d'anneaux à empiler et les cubes d'un jouet de cubes à empiler); celles-ci se sont avérées affectées par l'apparence visuelle des jouets. Plus spécifiquement, les mères produisaient une proportion beaucoup plus importante de références spécifiques (p. ex. l'anneau bleu) que de références non spécifiques (p. ex. ceci ou cela) lors du jeu avec des jouets dont l'apparence était simplifiée. En effet, la proportion de références spécifiques était trois fois plus élevée dans les situations de jeu avec des jouets dont l'apparence était simplifiée que dans les situations de jeu avec des jouets « chargés » sur le plan visuel. L'utilisation de jouets ayant une apparence « chargée » a également réduit le nombre total de références aux parties des jouets, ainsi que la fréquence d'exposition des enfants à des mots de vocabulaire permettant de décrire les caractéristiques des jouets (p. ex. les termes utilisés dans les références spécifiques pour décrire les couleurs). Ces résultats soulignent l'importance de considérer l'apparence visuelle, et ce, particulièrement dans les contextes où l'on cherche à encourager les échanges entre un adulte et un enfant, ainsi que dans les situations de jeu où l'on cherche à augmenter l'exposition d'un enfant à un vocabulaire riche en informations.

This study is the result of a personal experience of the first author. One day, while visiting with a family whose young toddler-age son had recently been diagnosed with severe autism, she was in the kitchen watching him play with a toy. The toy was bright, with neon-coloured pieces that made a sound when they were inserted into corresponding slots. As he was playing, she wanted to join in by saying something about one of the pieces, except she could not figure out what to say. Like the other pieces of the toy, the piece she was trying to talk about was curvy-shaped and its colour was a difficult-to-name shade of pink (a fuchsia-red-pink). As she struggled to describe both its colour and shape, and resorted to an unsatisfying "*that one*," his mom noticed and said, "*We call that the hammery-thing*." The irony of this situation was immediately apparent to her. Here was a child struggling with acquiring his first words and for whom it might have been especially beneficial to have a toy that made it easy for an adult to talk about its parts in a clear way. But the design and appearance of the parts of this particular toy made naming any part of this toy very difficult—indeed, almost impossible. As it happens, this toy was made by a popular toy brand and representative of newer toy designs that are often described on their packaging as "stimulating" or "educational." From this experience came the motivation for this study and the main question it pursued: Might a toy's visual design (e.g., unusual colours; large number of different, complex designs and patterns all visible at once) negatively impact the specificity, and thus clarity, of mothers' references to a toy's main parts?

Children develop physically, linguistically, cognitively, and socially through play, and play specifically allows a parent to scaffold the experience according to the child's developmental needs (Vygotsky, 1967). In particular, play has been argued to foster language development in at least four ways: Play can (a) require symbolic thinking, (b) involve social interaction (primarily with parents), (c) expose children to a large amount of language, and (d) keep children engaged in the learning process if it is a child-led activity (Weisberg, Zosh, Hirsh-Pasek, & Golinkoff, 2013). Within the past decade, researchers have begun to focus on examining the effect of toys' design features on parent–child talk during play with toys. Although this research is limited in scope, first studies have demonstrated that certain features of toys do indeed affect the quality of associated parent–child talk, as will be discussed next.

Considering that the phrase "batteries not included" is a common disclaimer on many current toys, it is not surprising that the majority of studies on this topic have focused on how electronic features of toys may influence the quality of parent–child talk interactions during play. In one direct comparison between the interactions afforded by electronic and traditional toys, Wooldridge and Shapka (2012) observed parent and child (16 to 24 months) dyads playing with both electronic and traditional toys in a controlled laboratory playroom. The toys included three traditional toys (i.e., a shape sorter, plastic animals, and a picture book) and an electronic version of each. The play sessions were recorded and coded using the 3-point rating scale Parents Interacting with Children: Checklist of Observations Linked to Outcomes (PICCOLO; Roggman, Cook, Innocenti, Jump Norman, & Christiansen, 2009). for parental affection, responsiveness, encouragement, and teaching. With the traditional set of toys, significantly higher levels of parental responsiveness, teaching, and encouragement were observed. The largest difference was seen for parental teaching: on average, the set of electronic toys resulted in parent teaching ratings that were over 2.5 times lower than for traditional toys. The authors attributed these results to a shift in the parents' conceptualization of their role from "supporting the child's play" to "letting the toys do the talking."

Similar reductions in the quality features of parent talk have been found in other studies (e.g., Radesky & Christakis, 2016; Sosa, 2015). Sosa (2015) recorded parent and child (10 to 16 months) dyads playing with a set of electronic toys (e.g., baby cellphone), traditional toys (e.g., shape sorter), and books (e.g., book of farm animals) in their homes over 3 days. The most profound differences were observed for the electronic toys versus the other two traditional toy sets. The results suggested a diminishing of the quality of the parent–child interaction. Namely, during play with the electronic toy set, parents produced fewer conversational turns, attentive responses, and fewer words—especially content-specific words. Books surpassed traditional toys with respect to these measures, but to a lesser degree than both books and traditional toys compared to electronic toys. In a brief overview of eight studies involving electronic toys up to December 2015, Radesky and Christakis (2016) concluded that electronic toys reduce parents' verbal and non-verbal contributions during play and that although electronic toys may engage the child, they disengage the parent.

In addition to findings regarding impacts on the overall quality of the parent–child interaction, technological affordances on toys have been found to usurp the original purpose of a toy. Zosh et al. (2015) compared parent and child (20 to 27 months) interactions when playing with electronic versus traditional shape sorters. Although similar amounts of talk occurred overall with the two types of shape sorters, the quality of the talk differed in significant ways. Parents playing with the electronic shape sorter

produced less spatial language (e.g., shape names, place referentials such as *here* and *there*, locations, directions) and talked more about the non-shape related features and functions of the toy (e.g., pushing its buttons). The authors concluded that the additional electronic features detracted from the toy's intended purpose of fostering a better understanding of spatial concepts. Similarly, a study of parent–child pretend play found less pretense to be produced and the interaction to be more parent-directed when playing with an electronic toy house (Bergen, Hutchinson, Nolan, & Weber, 2009).

In the realm of books, rather than toys, certain features of electronic books (e.g., button consoles) have also been shown to reduce story-related talk and increase behavioural directions from parents during reading, impacting children's storyline comprehension (Parish-Morris, Mahajan, Hirsh-Pasek, Golinkoff, & Collins, 2013). Even certain non-electronic features like pop-ups in traditional books have been shown to distract children from learning new words and remembering the storyline (Tare, Chiong, Ganea, & DeLoache, 2010). It should be noted, however, that studies directed specifically towards parent and child talk during the sharing and reading of e-books versus traditional books has moved towards identifying, in a more nuanced manner, the particular features of e-books that can serve to enhance or impair different outcomes including vocabulary learning, decontextualized talk, and comprehension of the story (e.g., Bus, Takacs, & Kegel, 2015; Guernsey & Levine, 2015).

Most of the research on more current toys and the impact of their "bells and whistles" has focused on electronic versions of toys and books. Our study, however, explores another potentially influential aspect of the design of these toys, namely their *visual design features*. Even apart from technological features, such as sounds produced, newer toys are often designed with lots of colour, texture, and pattern features to stimulate a child's play, learning, or imagination. For example, the *Whoozit® Tip Top Tower Block™* toy, used in the current study, is accompanied by the manufacturer's description:

> Young children gain new skills and talents almost overnight. During this phase of dynamic development, Manhattan Toy's research-proven Whoozit collection of toys stimulates fundamental learning skills. Our Whoozit is featured on our tip top tower blocks, a stacking set that's a puzzle, too. Each side features a pattern: numbers, characters, shapes and stars; plus rattling rings (http://www.amazon.com/Whoozit-Tip-Top-Tower-Block/dp/B00157D4UA).

Although there may be a growing movement towards simpler and more traditional toys underway (Hirsh-Pasek & Golinkoff, 2006; Hirsh-Pasek & Zosh, 2017), a stroll down any big toy store aisle will readily confirm a "more is more" marketing approach to many toys, especially those often labelled as educational. Providing many features to stimulate a child is often presented as a positive attribute of a given toy to potential purchasers. We will henceforth refer to a toy with many such stimulating visual design features as a *busy toy* and contrast this with a *simple toy*.

In the current study, we sought to explore the effect of the visual design of a busy toy versus a simple toy on one important linguistic feature of a parent's talk while playing with his or her toddler: the specificity of a parent's verbal references to the main parts of the toy. Specificity in verbal referencing is vital to the smooth flow of a dyad's mutual understanding and interaction during play. A child must be able to understand easily and clearly what a parent is referring to for vocabulary acquisition to proceed smoothly (for a review, see Trueswell et al., 2016). This will be especially the case when a toy has multiple parts and if the intended purpose of the toy is to do something with the parts in a specific sequence, such as stacking ring or nesting block toys. For example, if there are four blocks or rings that are part of a toy, being able to refer to each block or ring in a clear manner will be helpful to both the parent and the child. Specificity may be achieved in different ways, such as by appealing to a distinctive, specific feature or attribute (e.g., colour) unique to each part or the use of a more idiosyncratic, agreed-upon name (e.g., the lid) via a "referential pact" (e.g., Brennan & Clark, 1996; Matthews, Lieven, & Tomasello, 2010). By whatever means, the key is that specificity is possible given the visual design features of the parts of the toy. If specificity is not possible, then one is left with only the choice to use a very general means to refer to a part such as the use of a demonstrative like *this one* or *that one*. The latter option is not only potentially ambiguous, but also in the context of early parent–child talk, reduces the opportunity for exposure to informative vocabulary that will be necessarily incorporated into more specific references (e.g., colour and size terms such as *the blue block* or *the biggest one*). Thus, the inability to achieve specificity in referencing can be viewed as another way in which a toy's design could negatively impact the quality of parent–child talk.

In our study, we examined a parent's references with respect to the parts of two toys: (a) the rings of a stacking ring toy and (b) the blocks of a nesting block toy. Our manipulation consisted of the creation of a simple visual design version of each of these toys by adapting the busy commercial version of each by, for example, covering up

the original detailed patterned sides of a toy's parts with material of a single colour. We believe this is the first study to examine the effect of the physical visual design of toys on the quality of an aspect of parent talk during toy play. Given that the features of the toys we manipulated in this study all pertained to aspects of the toy's visual appearance (i.e., not other physical properties such as the shape of the parts), we have referred to these features as the *visual design* of the toys.

Our goal to assess whether a toy's visual design affects the specificity of an adult's references to parts of the toy led us to address the following main hypothesis in this study: A greater proportion of a parent's references to the parts of a toy will be more specific (e.g., *the blue block*) when a toy's visual design is simple than when it is busy. That is, a toy's simpler visual design features may result in a greater number and proportion of informative (i.e., specific) references to its parts (e.g., *the green one*) because the smaller number of unique and easier-to-name features of the parts make it easier to describe the different parts in a clear and unambiguous manner. In contrast, similar to how the electronic affordances of toys have been shown to negatively influence parents' language during play with their child, a toy with busier visual design features may result in a greater number and proportion of references to its parts that are less specific and less informative (e.g., *that ring*) because of the preponderance of difficult-to-name features, or features that repeat across parts, that make it difficult to describe the parts in a clear and unambiguous manner. In addition, because we expected this effect of toy design to operate within-parent, toy design was manipulated within-parent. We compared the proportion of specific references for simple versus busy toys for a single group of parents who all played with both a simple and busy toy with their child.

Method

Participants

Twenty-four children (12 girls, 12 boys) ranging in age from 19.4–28.5 months (M = 23.5 months, SD = 66 days; M_{girls} = 23.1 months, SD = 79 days, age range: 19.4–28.5 months; M_{boys} = 23.8 months, SD = 54 days, age range: 21.5–27 months) participated in this study. All children were accompanied by their mother. Data from an additional six participants was excluded because English was not spoken during the session (n = 2), the child had a speech delay (n = 1), or the child was unwilling to play with the toy at all (n = 3). Participants were recruited via advertisements in local community centres and from the existing database at the UW Centre for Child Studies at the University of Waterloo. Participants were mostly of middle class, Western and Eastern European descent as is representative of the region. All children were exposed to no more than 20% of a second language at home, as recorded by parent report at the time of scheduling their visit to the lab. All participants received a certificate and a book as compensation for their participation in the study. All the procedures of this study received ethics clearance (Approval ORE#14874) from the Human Research Ethics Committee of the University of Waterloo's Office of Research Ethics.

Materials

The toys used in this study consisted of two versions, simple and busy, of both a stacking ring and nesting block toy (four toys altogether). As will be described further below, the busy version of each toy was largely the unadapted commercial version of the toy, whereas the simple version was created by modifying the commercial version to reduce its busy features. Pictures of these four toys are shown in figures 1a–1d. Both toys were indicated on their boxes to be appropriate for children aged 12 months and older. The manipulation of toy design was within-mother: each mother played with her child with one simple and one busy toy (i.e., the two toys played with were the simple stacking and busy nesting toys OR the simple nesting and busy stacking toys).

Busy and simple stacking ring toys. The stacking ring toy was the *Nooboo Symphonic Stacker™* from Manhattan Toy and was 22 x 22 x 24 cm. Since it originally made a sound when a ring was placed on the post, the batteries were removed from this toy for the purpose of this study. The toy consisted of four flower-shaped, plush rings of increasing size, with the last ring being an enclosed topper piece. A solid colour of shiny, satin fabric lined the bottom side of each ring (i.e., yellow, green, orange, or pink) while the top of each ring was covered in several different fabrics of various colours (e.g., blue, purple, orange), patterns (e.g., polka dots, swirls, stripes), and textures (e.g., corduroy, satin, felt). Small ribbon tags of various colours were also attached around the side of each ring (see Figure 1a). This original commercial version of the toy was the busy version of the stacking ring toy. The simple version of this toy was created by covering the top (i.e., multi-coloured, textured, patterned) of each ring with felt of the corresponding solid colour of the bottom side of the ring (see Figure 1b).

Busy and simple nesting block toys. The nesting block toy was the *Whoozit® Tip Top Tower Block™* from Manhattan Toy and was 15 x 15 x 41 centimetres. The toy consisted of four soft, plush blocks of increasing size. Each

block could be inserted into each other or stacked on top of each other. The top, outside surface of each block was lined with different coloured checkerboard fabric. One outside panel of each of the four blocks had numbers and shapes/faces with different colours. A second outside panel had multi-coloured stars, while a third outside panel had a yellow pathway with stars that lined up to create a continuous pathway across the four blocks. A final outside panel of each block was covered with a multi-coloured, swirling pattern that resembled paint splotches on a canvas (see Figure 1c). The inside of each block, however, was a single colour. (Rattling rings were attached to each block and were removed for both busy and simple toy versions.) Thus, to create a simple version of this toy, all outside panels of each block were covered with a solid colour of felt matching the inside colour to produce one blue, yellow, green, and pink block. To retain some visual interest for children, the pathway of stars was replicated from the original version (see Figure 1d).

Figure 1a

Photo supplied by authors.

Commercial version of the *Nooboo Symphonic Stacker*™ (Manhattan Toy) and its stacking rings used as the *busy* stacking toy in this study.

Figure 1b

Photo supplied by authors.

Modified *Nooboo Symphonic Stacker*™ (Manhattan Toy) used as the *simple* stacking toy in this study. The tops of its stacking rings were changed to a solid colour to match their original solid bottom colour.

Figure 1c

Photo supplied by authors.

Commercial version of the *Whoozit® Tip Top Tower Blocks*™ (Manhattan Toy) and its nesting blocks used as the *busy* nesting toy in this study.

Figure 1d

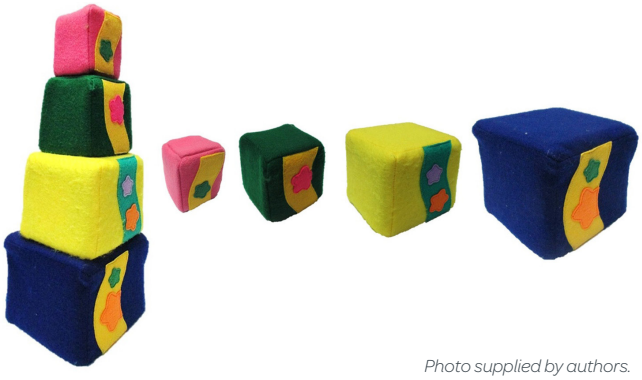


Photo supplied by authors.

Modified *Whoozit® Tip Top Tower Blocks*™ (Manhattan Toy) used as the *simple* nesting toy in this study. The outer sides of its blocks were changed to a solid colour to match their original solid inside colour.

Procedure

Parents and children were seated at a small table in the lab playroom. Parents were instructed that a small set of drawers labelled 1 and 2 contained the first and second toy to be played with. The drawer set kept the toys out of sight and reach of the child behind the parent. Ahead of time, one busy version of the nesting/stacking toy was placed into one drawer and one simple version of the other type of toy (stacking/nesting) in the other drawer according to fully counterbalanced orders within boys and girls.

Ahead of time, parents were given the general instruction to "play with each toy with your child as you would at home for as long as your child remains interested." When their child was no longer interested in the first toy, they were also instructed to return it to its bin and select the next toy from the second bin. The study's session ended when the second toy was returned to its bin. Every session was audio and video recorded for later transcription and analysis.

Transcription

All sessions were transcribed according to the CHILDES transcription system (MacWhinney, 2000, 2016). All speech from both the parent and child was transcribed, although in this study only a mother's references to a single part of the toy were the focus of analysis (see details of the coding scheme below). All transcripts were initially transcribed by the third author and reviewed by the fifth author at which time any discrepancies were discussed and adjusted accordingly. The second author conducted a final third review at the time of coding at which time no further discrepancies were noted.

Coding of Mothers' References to the Toy Parts

Identification and total number. To begin, all possible types of references to a single part of the toy (i.e., one of the four rings or top piece in the stacking ring toy; one of the four blocks in the nesting block toy) were identified in the transcripts by the second author and reviewed together with the first author. Thus, not included were (a) plural referents (e.g., *they, those*) that referred to more than one part at a time, (b) the use of *one* to mean the number one, (c) an utterance containing only the sole use of a colour or adjective term (e.g., one-word utterances such as *blue* or *spotty*), and (d) the pronoun *it* as it presumes the establishment of a commonly understood referent in contrast to, for example, the use of *this* or *that*. Also excluded were a few references to a single part using a label that was uttered within the context of pretending the part was something else (e.g., *a hat*) and accompanied by pretend actions such as placing the ring on top of the head. The reader will note, however, that if such a label was used outside of a pretend context (e.g., calling the top piece *a hat*) then these references were included.

Reliability coding with respect to this initial identification of references to a part of either of the two toys was carried out with 30% of the participants' transcripts by a research assistant blind to the hypothesis of the study and was found to be 100%. Thus, we felt confident that all possible ways in which mothers had referred to a part of either toy (e.g., *this/that one, the hat, the blue one*) had been captured. The CLAN program (MacWhinney, 2016) for use with CHILDES transcripts was used to confirm the frequency counts of all references to toy parts identified at the first stage. Overall, 552 references to a single part of a toy (308 for the stacking ring toy and 244 for the nesting blocks toy) were identified and, in a second stage of coding, subsequently classified by level of specificity, as described next.

Level of specificity. Each reference was categorized into one of three levels of specificity. Although we had initially anticipated a dichotomous *non-specific* (e.g., *that/this one*) versus *specific* classification (e.g., *the biggest one*), a further in-between category of *under-specific* was added to capture references that contained more information than the non-specific, but remained only semi-specific given multiple possible referents among the stacking rings or nesting blocks (e.g., *the big one*). Thus, the three levels of specificity were defined as follows for coding:

(a) *Non-specific.* The part of the toy was referred to solely by the use of the demonstrative pronouns *that/this (one/block/ring/box)*, the pronoun *one*, the determiners *another* or *other coupled* with *one* (e.g., *another one, other one*); the question forms *which/what one*?; two instances of the use of a pronoun (he/she); *one* accompanied by an evaluative adjective (e.g., *nice one*); or the word *next*. In all these cases, the mothers' utterance contained no information that could be used by the child to identify the part intended from any other part (i.e., *that one* could potentially apply to any of the four boxes or rings). As a result, these references were considered non-specific. Note that instead of *ring*, at times mothers also referred generally to all the rings using terms such as *flower* (due to their wavy shape) or *bracelet*.

(b) *Under-specific.* The part of the toy was referred to by using *one* or *flower/block/ring/box* accompanied by another term (e.g., *little*) that provided some information that could be potentially used to distinguish it from another part, but that was not unique to this part and could have referred to another part (or even other parts) as well. For example, the use of *big one* does not, in the

context of all the rings or blocks, uniquely describe any of the three rings or blocks that are bigger than the one smallest ring/block. Similarly, the reference *a green one* does not uniquely establish this part as being the only green one (cf *the green one* classified as specific below). Thus, these references remain under-specific, but nevertheless provide some exposure to new information and vocabulary than the wholly non-specific references. Indeed, the syntax of generic versus non-generic utterances is considered to be a form of linguistic input to which toddler-age children are sensitive (e.g., Gelman & Raman, 2003) and that is demonstrated in parental speech to children around this age (e.g., Gelman, Chesnick, & Waxman, 2005; Nyhout & O'Neill, 2014). As a result, this category was retained for analysis.

(c) *Specific.* The part of the toy was referred to via the use of a descriptor that could, with respect to that particular toy and part, uniquely and clearly distinguish it from all the other parts (e.g., *the yellow one*, with respect to a part in either one of the simple toys; *the biggest flower*, with respect to either the busy or simple stacking ring toy). Also included in this category were names given—albeit infrequently—to a part that was uniquely descriptive, such as a parent using *the lid* to describe the top-most part of the stacking rings.

In some instances, a single utterance referring to one part contained two levels of specificity, in which case we coded the utterance for the highest level (e.g., *that's the yellow one* was coded as specific rather than non-specific; *that's the big one* was coded as under-specific rather than non-specific). If an utterance contained a reference to two parts, each part received a code according to the descriptions above (e.g., *put the yellow one in the blue one* was coded as containing two specific references).

Reliability of this coding scheme was carried out in full for all 552 references by a Master's level student blind to the hypothesis of the study and was found to be excellent (99.3%, only three instances of a discrepancy).

Time played with toy. The length of the parent–child play session with each of the two toys was calculated. A play session was defined as beginning when the toy was first in joint view of the parent and child and ending when the parent returned the toy to its bin.

Results

Analysis of Time Played With Toys

Children and parents played with each toy for an average of 3 minutes and 59 seconds (SD = 2 minutes and 6 seconds). On average, the dyads played one minute longer with the simple version of the toys than the busy version of the toys. A paired samples t test, however, revealed that this difference in time played between the simple toys (M = 389.25 seconds, SD = 194.57 seconds) and the busy toys (M = 329.38 seconds, SD = 218.20 seconds) was not significant, $t(23) = 1.42$, $p = .168$.

Raw Frequency of Mothers' Production of the Three Reference Types for Each Toy Design

Table 1 displays the raw frequency of each type of reference per toy design over all 24 mothers. From **Table 1** it can be seen that, consistent with our main hypothesis, of all 148 specific references observed, 82.4% occurred during play with a simple toy as opposed to a busy toy (17.6%). It is also noteworthy that 62.5% of all 552 observed references to a toy's part (collapsed over specific, under-specific, and non-specific) were produced during play with the simple toy, as compared to only 37.5% with the busy toy. A paired samples t test confirmed that the difference in mean overall frequency of referencing between the simple and busy toy designs (M_{Simple} = 14.38, SD = 7.25 vs. M_{Busy} = 8.63, SD = 8.65) was significant, $t(23) = 2.71$, $p = .013$.

Table 1

Total Number of Each of the Three Reference Types Produced Across all Mothers (n = 24) who Each Played With a Simple and a Busy Toy

	Toy Design		
Reference Type	Simple	Busy	Total
Specific	122	26	148
All non- & under-specific	223	181	404
Under-specific	76	66	142
Non-specific	147	115	262
Total	345	207	552

Proportion of Mothers' References of Each Reference Type for Each Toy Design

A proportion score for each of the three types of references was calculated for each mother separately out of the total number of her references while playing with the simple toy, and also separately out of the total

number of her references while playing with the busy toy. Mean proportion scores are shown in **Table 2**. No mother produced zero instances of referencing while playing with the simple toy. However, two mothers produced no references of any type when playing with the busy toy and were therefore excluded in our ANOVA analyses with proportion data as described below. Thus, the proportions shown in **Table 2** are based on a final sample size of 22.

Table 2

Proportion of Mothers' References at Each Reference Level out of the Total (per Mother) per Toy Design Type

	Toy Design	
Reference Type	Simple *M (SD)*	Busy *M (SD)*
Specific	.360 (.277)	.076 (.129)
All non- & under-specific	.640 (.277)	.924 (.129)
Under-specific	.209 (.179)	.226 (.182)
Non-specific	.431 (.230)	.698 (.236)

Note. All these proportions are based on the final sample size of 22 used in our ANOVAs given that two mothers produced 0 references of any type with the busy toy.

Analyses of Proportions

We conducted three omnibus repeated-measures ANOVAs (n = 22) with proportion of specific, non-specific, or under-specific references as the dependent measure; Sex (boy or girl) and Toy Pair (e.g., whether a dyad received as a pair of toys the busy stacking rings and simple nesting blocks or the simple stacking rings and busy nesting blocks) as the between-subjects variables; and Toy Design (simple or busy) as a within-subject variable. A more conservative alpha value of α = .017 was adopted to take into account the three ANOVAs conducted (α = .05/3).

Did mothers produce a significantly greater proportion of references classified as specific when playing with the simple toy compared to the busy toy? Supporting our hypothesis, there was a significant main effect of toy design, $F(1, 18) = 22.01$, $p < .001$, $\eta_p^2 = .550$. On average (see **Table 2**), the proportion of mothers' references to a toy's parts that were specific was significantly greater when playing with the simple toy (36%) than the busy toy (7.6%). No significant main effects of toy pair or sex, or any significant interactions, were revealed in either this analysis or the further two ANOVAs of under-specific and non-specific references (p-values = .117 to .985). As a result, our discussion will focus on the effect of toy design.

When the results were examined at the individual level, it was clear that this pattern held for the majority of mothers when they played with both toys with their child. Namely, 77% (n = 17) of the 22 mothers produced a greater proportion of specific references with the simple toy than with the busy toy, 18% (n = 4) produced no specific references for either the simple or busy toy, and only 5% (n = 1) produced a lower proportion of specific references with the simple toy than with the busy toy. If all 24 mothers are considered, these percentages are 79%, 17%, and 4%, respectively.

From the transcripts, with the goal of looking at the content of these specific references, it was found that with the simple toys, 77% (n = 94) of all of the mothers' 122 specific references relayed information about the colour of the part (e.g., *the blue flower*). Next, in order of frequency, mothers provided unique labels for the part (19%, n = 23; e.g., *the lid, the top*) and information about size (4%, n = 5; e.g., *the biggest one*).

In contrast, when playing with the busy toy, of the 26 specific references produced by only seven mothers overall, information was specified most often in the form of size (42%, n = 11) or via a unique label (35%, n = 9). Next most often, mothers specified colour (15%, n = 4) and pattern (8%, n = 2; e.g., *the one with polka dots*).

Proportion of non-specific references with simple versus busy toys. Our analysis revealed a significant difference between simple and busy toys for the proportion of non-specific references, $F(1, 18) = 20.01$, $p < .001$, $\eta_p^2 = .526$, but in the opposite direction to that of specific references. On average (see **Table 2**), the proportion of mothers' references to a toy's parts that were non-specific was greater when playing with the busy toy (69.8%) than the simple toy (43.1%). Mothers produced few to no specific references when playing with busy toys and thus non-specific references represented a much greater proportion of all references for busy toys.

Under-specific references to the toy parts. Under-specific references represented about one fifth of all references. On average, the proportion of mothers' references to a toy's parts that were under-specific did not differ significantly for simple (20.9%) or busy (22.6%) toys,

$F(1, 18) = .127$, $p = .726$, $\eta_p^2 = .007$.

Discussion

Our results demonstrate that a toy's visual design—simple or busy—affects the specificity, and thus the clarity, of mothers' references to the main parts (rings, blocks) of a stacking ring and nesting block toy they used when playing with their toddler. Consistent with our original hypotheses, when considering references at a at a non-specific, under-specific and specific level, a significantly greater number and proportion of mothers' references were specific (e.g., *the green ring; the biggest one*) when the dyad played with a toy with a modified simple visual design than when they played with a toy with the original (commercial) busy visual design (see toys in figures 1a-d). The difference in reference specificity observed between toy designs with respect to specific references was not subtle. The mean proportion of mothers' specific references dropped from a maximum of over one third (36%) when playing with the simple toy to only 7.6% when playing with the busy toy. Or stated conversely, the mean proportion of references produced with a simple toy that were specific was more than four times larger than the proportion produced by the same mothers playing with a busy toy.

Of the two remaining categories of references, non-specific references represented a significantly greater proportion of all references for busy toys (69.8%) than simple toys (43.1%). A middle category of under-specific references (e.g., *the big one; the next one*), representing about one fifth of total references, was not impacted by a toy's visual design. These results overall are not attributable to differences between mothers given the within-participant manipulation of toy design (simple vs. busy).

Looking at the raw data shown in **Table 1**, one can see that specific references represented at most about one third of all references (35%, 122/345) for the simple toy, but that this decreased to only one eighth of all references (12.5%, 26/207) for the busy toy. Another interesting finding from our study looking at the raw data is that of all 552 observed references, 62.5% were produced when playing with the simple toy. Thus, it appears that the busy version of the toys not only made it more difficult for mothers to produce clear (specific) references to the parts of these toys from the analyses above, but also significantly reduced mothers' overall propensity to reference the main parts of these toys.

Further, our results suggest that a toy's visual design also impacted the vocabulary children were exposed to via mothers' referencing. That is, when playing with their mother with a simple visual design toy they were exposed much more frequently to vocabulary about the features that could distinguish the parts of the toy. In the case of the simple toys, this overwhelmingly took the form of colour terms (e.g., *the blue block, the orange one, the green ring*). Indeed as described previously, within the category of specific references for simple toys, there were 94 instances of the use of colour terms. For busy toys, however, specific references were infrequent and there were only four instances of the use of a colour term among them as references generally took the form *this/that one*. Given that the age of the children in this study (i.e., 23 months) places them firmly within the stage of language acquisition where vocabulary is growing rapidly, this negative impact of a toy's busy visual design on children's exposure to more informative vocabulary, such as colour terms, should be noted.

Children's word learning was not explored in this study, but it is uncontroversial to state that to learn new vocabulary and make distinctions among similar terms, such as colour or size terms, it is advantageous if children can encounter these terms more frequently in different settings. In this study, the busy visual design of two toys significantly reduced the amount of information mothers' provided when referencing these toys' parts compared to modified, simpler visual design versions of the same two toys. Empirical studies have clearly shown that the diversity of vocabulary input to children (of the same age as children in this study) is positively correlated with children's later vocabulary diversity, even with quantity controlled (e.g., Hart & Risley, 1995; Rowe, 2012). Thus, how easily a toy's visual design affords opportunities for parents to use diverse and informative vocabulary would appear to be a feature of a toy for parents, educators, and early intervention and speech-language professionals to consider.

There are indeed many toys in the marketplace beyond the two used in this study where such a consideration is relevant. For example, Fisher-Price has introduced a new version of their *Brilliant Basics Rock-A-Stack™, the Rock-A-Stack Pink Stacking Rings™*. Instead of the original classic blue, green, yellow, orange, and red stacking rings, the pink version features one blue base ring topped by four rings in successively lighter shades of pink (see **Figure 2**). For any adult, the pink version is likely to pose a greater challenge with respect to finding a way to refer clearly to one of the four different-shades-of-pink rings, especially using language that would be age-appropriate and easily understood by a toddler.

Figure 2

Photo supplied by authors.

Brilliant Basics Rock-A-Stack™ and *Rock-A-Stack Pink Stacking Rings™* by Fisher-Price.

It is most likely clear to readers from **Figures 1b** and 1d how easy it was for a parent to refer to each part of the simple toys clearly by, for example, using colour terms. What may be less obvious to glean from **Figures 1a** and **1c** is exactly how difficult it was for parents to find a way to refer to parts of the busy toys and the lengths parents went to in order to try to attain a greater level of specificity. For example, for the four busy stacking rings, a blue colour is shared by all rings, two depict stripes, one depicts spirals (not a frequent toddler-age vocabulary term), and one depicts something almost like polka dots but the dots are egg-shaped. For the four busy nesting blocks, similar colours are found on all blocks as well as difficult-to-name patterns (e.g., checkerboard). Other panels have difficult-to-name features such as one panel that depicts the numbers 1–4 but the corresponding pictures include items not easy to name including the *Whoozit*® face used on other toys in the line. As a result, we observed some mothers even try to introduce completely different dimensions by which to try to distinguish one part clearly (e.g., *the daddy*). With the busy nesting blocks, we also observed mothers talk about features on the panels of the toy and struggle to name them (e.g., *funny guy, smiley face, fireworks*).

Interestingly, we think, the effects of the toys' differing visual design in this study appeared to produce effects on the quality of mothers' references in terms of their level of specificity in a similar manner to how electronic features of toys have been shown to reduce parents' language related to the function of the toy (e.g., to highlight spatial language with a shape sorter; Zosh et al., 2015). Although in the literature on electronic toys some have argued their features result in parent disengagement (i.e., the child plays largely alone), we did not observe a similar disengagement with busy toys in our study. Rather, we would argue that the busy toy led the same mother who used specific references to toy parts when playing with the simple toy to be much less effective at doing so with the busy toy. As a result, the child playing with the busy toy with his or her mother was exposed, proportionally, to much more talk that simply referenced the parts of the toy as *this one* or *that one* and very little to no opportunity to hear more specific references used that made identification of the intended part much clearer (e.g., *the orange flower*).

In effect, the referencing occurring with busy toys seemed counter-productive to some of the main play functions of the toys that children and parents might be

trying to engage in together such as identifying, one-by-one, the different parts of the toy in a particular order to reproduce the original stack of rings, create a tower of nested blocks, nest the blocks within each other, or line up the parts by size. Indeed, the fact that the toys with a busy visual design reduced specific references to the toy parts so significantly, we argue, could be viewed as limiting parents' ability to provide scaffolding support to their child while playing with the toy. As a result, the type of adult linguistic input and engagement viewed as essential to explaining how children are successfully exposed to enriching language experiences during play (Weisberg et al., 2013) was hindered by the busy visual design of toys. Moreover, responding contingently to a child's actions has been shown to be an important aspect of enriching, instructive play (Fisher, Hirsh-Pasek, Newcombe, & Golinkoff, 2013; Weisberg et al., 2013). Thus, the ability of parents to easily be able to talk about pieces of a toy while playing with their child—rather than searching for a description as the first author found herself doing in the introductory anecdote to this study—would serve to increase the possible opportunities for contingent responses.

Our results also underscore the importance of considering situational and contextual factors when examining parent–child talk. It would not be appropriate to conclude from our results that some mothers had more or less informative or specific styles of referencing overall, or that the results are due to the greater or lesser talkativeness of some mothers. Rather, a mother's referencing was impacted by the visual design of the toy (simple or busy) and changed depending on which toy she was playing with together with her child. The toy's design is impacting the ability of mothers to refer clearly, or not, to the toys' parts. In this sense, our results are similar to findings that the complexity of mothers' talk varies as a function of book genre (e.g., vocabulary flashcard type picture book versus a story picture book; Nyhout & O'Neill, 2013).

The current study employed only two toys and focused on the impact of visual design on the specificity of a parent's references to parts of the toy. Potentially valuable extensions of this research could include an examination of the impact of visual design on other aspects of parent–child talk. For example, perhaps there exist shape-sorters with overly busy visual designs, or shapes in non-traditional styles, that detract from spatial language used by a parent in the same way that electronic affordances have been found to misdirect a parent's focus (Zosh et al., 2015). It is even possible that some toys would impact parents' talk for both reasons: the set of stacking rings used in our study, for example, also had a sound feature that we turned off (i.e., a sound occurred when rings were stacked on the post). Another avenue for further exploration could be a consideration of visual design effects with respect to toys for older children where the impact on adult–child, child–adult, and peer-to-peer talk might be of interest.

Finally, we note that both of the commercial toys we used in this study are well-regarded toys. The *NooBoo Symphonic Stacker™* received the Oppenheim Toy Portfolio Platinum Award 2006, the National Parenting Publications Awards 2006 Gold Award Infant/Toddler, and was listed on the National Association for Gifted Children Holiday Educational Toy List 2006. Toys among the *Whoozit®* collection have also won awards, including the Oppenheim Toy Portfolio Gold Seal award (https://www.amazon.com/Whoozit-Tip-Top-Tower-Block/dp/B00157D4UA).

It is not clear whether evaluations of these toys pertain largely to contexts in which a child is playing alone with them or together with an adult. Our results do not (and cannot) speak to the value of these toys in a solitary play context. That would require a different study with different measures (e.g., children's sustained attention). What our results speak to is the potential impact of busy visual designs on one aspect of the quality of parents' talk—the specificity of their references to parts of the toy—when playing with the toy together with their child. Here our results are clear: busy visual designs reduced the frequency and proportion of mothers' informative references and led to a preponderance of non-specific references.

Conclusion

This is one of the first studies of the impact of the visual design of toys on the quality of parent talk. Our results would support a "less is more" approach. However, just as has happened with further research exploring the positive and negative aspects of electronic versus traditional picture books, the answer is unlikely to be so simple. Instead, for toys, just as for e-books, it may be a case of understanding and exploring at a more subtle and specific level how certain features impact different aspects of play and talk with the toy in a negative or positive way. We believe our study begins a discussion of the potential ways in which a toy's visual design can impact the quality of parent–child talk as they play with the toy together and highlights the importance of considering a toy's visual design especially in contexts where the goal may be to foster adult–child language and a child's exposure to more information-rich vocabulary terms during toy play with an adult.

References

Bergen, D., Hutchinson, K., Nolan, J. T., & Weber, D. (2009). Effects of infant–parent play with a technology-enhanced toy: Affordance-related actions and communicative interactions. *Journal of Research in Childhood Education, 24,* 1–17. doi:10.1080/02568540903439342

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Memory, Learning, and Cognition, 22*, 1482–1493. doi:10.1037/0278-7393.22.6.1482

Bus, A. G., Takacs, Z. K., & Kegel, C. A. T. (2015). Affordances and limitations of electronic storybooks for young children's emergent literacy. *Developmental Review, 35*, 79–97. doi:10.1016/j.dr.2014.12.004

Fisher, K. R., Hirsh-Pasek, K., Newcombe, N., & Golinkoff, R. M. (2013). Taking shape: Supporting preschoolers' acquisition of geometric knowledge through guided play. *Child Development, 84,* 1872–1878. doi:10.1111/cdev.12091

Gelman, S. A., Chesnick, R., & Waxman, S. R. (2005). Mother–child conversations about pictures and objects: Referring to categories and individuals. *Child Development, 76,* 1129–1143. doi:10.1111/j.1467-8624.2005.00876.x-i1

Gelman, S. A., & Raman, L. (2003). Preschool children use linguistic form class and pragmatic cues to interpret generics. *Child Development, 74,* 308–325. doi:10.1111/1467-8624.00537

Guernsey, L., & Levine, M. H. (2015). *Tap, click, read: Growing readers in a world of screens*. San Francisco, CA: Jossey-Bass.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MA: Brookes.

Hirsh-Pasek, K. A., & Golinkoff, R. M. (2006). Choose toys. In S. Ettus (Ed.), *The experts' guide to the baby years: 100 things every parent should know* (pp. 246–249). New York, NY: Random House.

Hirsh-Pasek, K. A., & Zosh, J. M. (July 24, 2017). *Toys are tools: Minds make the magic.* [Blog Post]. Retrieved from http://www.thegeniusofplay.org/genius/expert-advice/articles/toys-are-tools-minds-make-the-magic#.WudIJIIpBBw

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Erlbaum.

MacWhinney, B. (2016). *The TalkBank project: Tools for analyzing talk – electronic edition: Part 2: the CLAN programs*. Retrieved from: http://childes.talkbank.org/manuals/clan.pdf.

Matthews, D., Lieven, E., & Tomasello, M. (2010). What's in a manner of speaking? Children's sensitivity to partner-specific referential precedents. *Developmental Psychology, 46,* 749–760. doi:10.1037/a0019657

Nyhout, A., & O'Neill, D. K. (2013). Mothers' complex talk when sharing books with their toddlers: Book genre matters. *First Language, 33*, 115–131. doi:10.1177/0142723713479438

Nyhout, A., & O'Neill, D. K. (2014). Storybooks aren't just for fun: Narrative and non-narrative picture books foster equal amounts of generic language during mother–toddler book sharing. *Frontiers in Psychology, 5,* 325. doi:10.3389/fpsyg.2014.00325

Parish-Morris, J., Mahajan, N., Hirsh-Pasek, K., Golinkoff, R. M., & Collins, M. F. (2013). Once upon a time: Parent–child dialogue and storybook reading in the electronic era. *Mind, Brain, and Education, 7,* 200–211. doi:10.1111/mbe.12028

Radesky, J. S., & Christakis, D. A. (2016). Keeping children's attention: The problem with bells and whistles. *JAMA Pediatrics, 170,* 112–113. doi:10.1001/jamapediatrics.2015.3877

Roggman, L. A., Cook, G. A., Innocenti, M. S., Jump Norman, V. K., & Christiansen, K. (2009). *PICCOLO (Parenting interactions with children: Checklist of observations linked to outcomes).* Logan, UT: Utah State University.

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development, 83,* 1762–1774. doi:10.1111/j.1467-8624.2012.01805.x

Sosa, A. V. (2015). Association of the type of toy used during play with the quantity and quality of parent–infant communication. *JAMA Pediatrics, 170,* 132–137. doi:10.1001/jamapediatrics.2015.3753

Tare, M., Chiong, C., Ganea, P., & DeLoache, J. (2010). Less is more: How manipulative features affect children's learning from picture books. *Journal of Applied Developmental Psychology, 31,* 395–400. doi:10.1016/j.appdev.2010.06.005

Trueswell, J. C., Lin, Y., Armstrong, B., Cartmill, E. A., Goldin-Meadow, S., & Gleitman, L. R. (2016). Perceiving referential intent: Dynamics of reference in natural parent–child interactions. *Cognition, 148,* 117–135. doi:10.1016/j.cognition.2015.11.002

Vygotsky, L. S. (1967). Play and its role in the mental development of the child. *Soviet Psychology, 5*(3), 6–18. doi:10.2753/RPO1061-040505036

Weisberg, D. S., Zosh, J. M., Hirsh-Pasek, K., & Golinkoff, R. M. (2013). Talking it up: Play, language development, and the role of adult support. *American Journal of Play, 6,* 39–54.

Wooldridge, M. B., & Shapka, J. (2012). Playing with technology: Mother–toddler interaction scores lower during play with electronic toys. *Journal of Applied Developmental Psychology, 33,* 211–218. doi:10.1016/j.appdev.2012.05.005

Zosh, J. M., Verdine, B. N., Filipowicz, A., Golinkoff, R. M., Hirsh-Pasek, K., & Newcombe, N. S. (2015). Talking shape: Parental language with electronic versus traditional shape sorters. *Mind, Brain, and Education, 9,* 136–144. doi:10.1111/mbe.12082

Authors' Note

Current author affiliations are T. Deglint: HearSay Speech and Hearing Centre and St. Joseph's Health Centre, Guelph; A. McKinnon: Swanson, Moss, Heimpel & Associates, Waterloo; A. Nyhout: University of Toronto/OISE, Toronto; and J. Scott: Waterloo Region District School Board, Waterloo.

Correspondence concerning this article should be addressed to Daniela O'Neill, Department of Psychology, University of Waterloo, Waterloo, ON N2L 3G1 Canada. Email: doneill@uwaterloo.ca

Acknowledgements

We thank the parents and children of the Kitchener-Waterloo region for their participation in this study. This work was supported by a Social Sciences and Humanities Research Council (Research Development Initiatives) grant (2009-2011) and Robert Harding Humanities and Social Sciences Award (2013) awarded to the first author.

Disclosures

No conflicts of interest, financial or otherwise, are declared by the authors.

Assessment, Diagnosis, and Recovery From Language Disorder at Kindergarten Age: Research Review and Clinical Discussion

Évaluation, diagnostic et récupération du trouble du langage chez les enfants en âge de commencer la maternelle : revue de la littérature et discussion clinique

KEYWORDS

LANGUAGE DEVELOPMENT
LANGUAGE ASSESSMENT
DIAGNOSIS
LANGUAGE DELAY
LANGUAGE IMPAIRMENT
LANGUAGE DISORDER
PRESCHOOL
ILLUSORY RECOVERY
DEVELOPMENTAL LANGUAGE DISORDER
SPECIFIC LANGUAGE IMPAIRMENT

Monique Charest
Paige Borger
Carolyne Chan
Kaitlin Sanders
Beatrice Yip
Lu-Anne McFarlane
Phyllis Schneider

Abstract

The research literature reveals two seemingly contradictory findings about the trajectory of developmental language disorders identified in the preschool years. Some studies suggest that many children achieve normal language by about the age of kindergarten entry. Other studies, however, indicate that most language disorders persist over a much longer time frame. Scarborough and Dobrich (1990) suggested that the apparent inconsistency in findings might arise as the result of patterns of illusory recovery at around kindergarten age. Periods of plateau in typical language development may allow children with language disorders to appear to have caught up to their peers, only to experience renewed challenges in the school years. The purpose of the current paper was to examine the literature for evidence in support of or inconsistent with the illusory recovery hypothesis and to discuss the clinical implications that follow from the evidence.

Monique Charest, Paige Borger, Carolyne Chan, Kaitlin Sanders, Beatrice Yip, Lu-Anne McFarlane, and Phyllis Schneider

University of Alberta, Edmonton, AB, CANADA

Abrégé

On retrouve, dans la littérature, deux résultats apparemment différents quant à la trajectoire des enfants ayant un trouble développemental du langage identifié pendant la période préscolaire. En effet, les résultats de certaines études suggèrent que de nombreux enfants vont présenter des habiletés langagières dans les limites de la normale lorsque ceux-ci seront en âge de commencer la maternelle. Cependant, les résultats d'autres études suggèrent que la plupart des enfants vont présenter un trouble du langage qui va persister sur une période beaucoup plus longue. Scarborough et Dobrich (1990) ont suggéré que ces résultats apparemment différents pourraient découler d'un patron de récupération illusoire (*illusory recovery*) qui se produirait lorsque les enfants sont en âge de commencer la maternelle. Les périodes où les enfants au développement typique atteignent un plateau dans le développement de leurs habiletés langagières pourraient permettre aux enfants ayant un trouble du langage de donner l'impression de rattraper leurs pairs sur le plan du langage, pour néanmoins connaître de nouveaux défis après l'entrée à l'école. Le présent article avait pour objectif d'investiguer la littérature afin d'identifier les sources de support, ou encore, les sources inconsistantes avec l'hypothèse de récupération illusoire, en plus de discuter des implications cliniques qui en découlent.

An important part of the work done by speech-language pathologists (S-LPs) in early childhood settings is the assessment and identification of children in need of support for language development. S-LPs use the information obtained from assessments to advocate for their clients' access to timely and appropriate services and to support children to maximize their potential. In many jurisdictions, the time spanning entry to kindergarten and Grade 1 involves transitions in educational setting, service provision, and/or funding source. Children previously identified as having a language delay, impairment, or disorder—acknowledging that different clinicians may adopt different terms—are often reassessed during this transition period. The conclusions that follow from these assessments have consequences for the nature and extent of supports that children receive as formal schooling begins—or even whether supports are identified as needed at all. This paper will discuss challenges and considerations for assessment and diagnostic decisions at kindergarten age, with a particular focus on implications when assessment results suggest that a child's language disorder has resolved.

This review and discussion is motivated by a well-known paradox in the literature on children with early-identified language disorders. While some research indicates that a substantial proportion of children will "recover" or achieve normal language status by about kindergarten age (Bishop & Edmundson, 1987; LaParo, Justice, Skibbe, & Pianta, 2004), other research indicates that language disorders tend to persist for much longer (Aram, Ekelman, & Nation, 1984; Stark et al., 1984; see Nippold & Schwarz, 2002, for further discussion). Several decades ago, Scarborough and Dobrich (1990) identified a possible explanation for this apparent paradox: They suggested that much of the recovery seen around age 5 may in fact be illusory. They argued that such a situation could arise as a result of the non-linear nature of language growth in typical development, characterized by alternating periods of growth and plateau in skills. When typical development plateaus, children who are following a slower course of development may appear to catch up, only to be left behind when typical language development once again accelerates. This characterization may be particularly fitting around kindergarten age as this is a time of some transition with respect to language achievements and needs. In typical development, many of the building blocks of language have largely been mastered, such as grammatical morphemes and control of most of the sentence structures of the language, including both simple sentences (e.g., *Michael was crying*) and complex sentences (e.g., *Michael was crying because he dropped his ice cream*).

During the school years, however, language is increasingly used as a tool for learning, requiring increasingly sophisticated content, form, and use (e.g., Pence-Turnbull & Justice, 2012). The development of reading, writing, and using written language as a basis for learning also requires extension of language skills. The prospect of kindergarten-age illusory recovery presents the risk that some children will be prematurely identified as no longer in need of language supports right at the time that they are transitioning to the more demanding context of formal schooling. As a result, they will potentially miss out on crucial years of support and/or have academic, social, or other challenges be misunderstood.

In support of the idea that challenges may disappear around kindergarten age only to reappear later, Scarborough and Dobrich (1990) presented the data of four children from a longitudinal sample who were retrospectively identified with significant expressive language delays at 30 months based on the absence of word combinations at that age. By 5 years of age, they became essentially indistinguishable from a control group of children on measures of mean length of utterance (MLU), lexical diversity, grammatical complexity (as measured by the Index of Productive Syntax; Scarborough, 1990), and pronunciation accuracy. These children were considered to have typical language at age 5 according to the measures employed in that study. Data from the control children showed plateaus in these language measures over the course of the preschool years. By Grade 2, three of the four children with a history of language delay presented with severe reading disabilities. A similar pattern of low oral language scores showing normalization at age 5 was reported for a larger group of children who were later identified with dyslexia (Scarborough, 1991).

The data that Scarborough and Dobrich (1990) and Scarborough (1991) reported focused on reading outcomes, and indeed there is much evidence that language disorders are associated with elevated risk for later difficulties with reading and writing, stemming both from challenges with decoding and challenges with comprehension of what has been read (Botting, 2007; Catts, Fey, Tomblin, & Zhang, 2002). The illusory recovery hypothesis is not specific to later reading outcomes—it can be applied to both oral and written language. The primary focus of the current review is on oral language outcomes, although where appropriate both are reported. The concept of illusory recovery raises

several important questions: How frequent or likely is recovery from preschool-identified language disorders? Is such recovery typically maintained when assessments are conducted at later ages? Does the risk of false or apparent recovery apply across language broadly, or is it dependent on how language is measured? The following sections will review evidence regarding recovery and persistence of language disorders before turning to consider language measures that may be sensitive to language needs and risk at kindergarten age.

There is evidence that the likelihood of recovery is lower when concerns extend to non-verbal cognition (Bishop & Edmundson, 1987). The review that follows thus focuses on children with language difficulties but with no identified impairments to non-verbal cognition. There are a number of different terms that have been used over the years to refer to children who fit this general profile. The most common term used in recent decades for research purposes is *specific language impairment*, although it is used much less often clinically (Bishop, 2014). More recently, a consortium of experts has proposed adopting the term *developmental language disorder* (DLD) to refer to children with unexplained language difficulties (Bishop, Snowling, Thompson, Greenhalgh, & the CATALISE-2 consortium, 2017). The sections that follow use DLD, except where the term *delay* is arguably appropriate, such as in reference to late talkers. For excellent overviews of the issues surrounding terminology, readers are referred to Bishop (2014) and Bishop et al. (2017).

Trajectories and Outcomes of Developmental Language Disorder

Several seminal studies of language outcomes in DLD concluded that DLD identified in childhood is often persistent, with language-based difficulties lasting into the school years and beyond (Aram et al., 1984; Stark et al., 1984; see Nippold & Schwarz, 2002, for further discussion). For example, Aram et al. (1984) reported on the 10-year outcomes of children who had originally been identified as having language disorders between the ages of 3;5 (years; months, the youngest child at initial assessment) and 6;11 (the oldest child at initial assessment). The children had been diagnosed with a language disorder by a certified S-LP and were all receiving some form of intervention. At ages 13–16, the primary language outcome measure was the Test of Adolescent Language (Hammill, Brown, Larsen, & Wiederholt, 1980). Of the 16 participants who had broadly normal-range nonverbal IQs (all with full scale IQs above 70), 13 scored well below the average range on the Test of Adolescent Language (Hammill et al., 1980) composite: 10 scored more than 2 standard deviations below the mean, and three scored between 1.75 and 2 standard deviations below the mean. The remaining three children obtained composite *z* scores of -0.73, -0.67, and 1.13. Similarly, Stark et al. (1984) examined outcomes at ages 8 to 12 of children originally tested at ages 4.5 to 8 and reported that 22 out of 29 children with DLD still met the study's clinical criterion for language disorder (a criterion based on discrepancy between an estimated "mental performance age" and language age estimates, considered appropriate at the time).

The conclusion that language disorders are persistent also holds in more recent studies that examined trajectories from kindergarten age onward. Studies show that the majority of children with language disorders documented at age 5 or later can be expected to show language-based difficulties throughout the school years or into adulthood (Beitchman et al., 1994; Botting, Faragher, Simkin, Knox, & Conti-Ramsden, 2001; Johnson et al., 1999; Stothard, Snowling, Bishop, Chipchase, & Kaplan, 1998; Tomblin, Zhang, Buckwalter, & O'Brien, 2003).

In research examining younger age groups with outcomes measured at around kindergarten age, a somewhat different picture emerges. In studies that have focused on late talkers identified at age 2, the general picture that emerges is one in which the majority of children move into the typical range by about age 5 (Paul, 1996; Whitehurst & Fischel, 1994; see Paul & Roth, 2011, for further discussion). However, as a group, the children continue to perform at the lower end of the typical range through the school years and obtain significantly lower scores on language and literacy measures relative to age peers matched for socioeconomic status (Rescorla, 2002, 2005).

In studies of children who have identified language disorders at 3 and 4 years, yet another picture emerges. The proportion of those that score within the normal range on language assessments as they approach kindergarten is somewhat lower than the proportion reported for late talkers, yet substantial nonetheless. In three studies published almost 30 years apart, approximately 45% of children identified with DLD as preschoolers were considered to have typical language skills, according to the studies' different criteria, when assessed around kindergarten age. These proportions were seen in research following children from 4 to 5.5 years (Bishop & Edmundson, 1987), 3 to 4.5 years (LaParo et al., 2004), and 4 to 5 years (Eadie et al., 2014). In the Bishop and Edmundson (1987) study, the participants were referred by pediatricians and S-LPs, meaning that

they came to the study with clinically-identified language concerns or diagnoses. In the LaParo et al. (2004) and Eadie et al. (2014) studies, children categorized as having DLD were identified from a larger cohort on the basis of a test score. Although it may seem questionable to apply the label of DLD on the basis of test scores alone, in the absence of previously identified concerns or functional observations, the results of these studies taken together nonetheless provide some insight into the stability of low language scores.

On the surface, the data seem to point to positive kindergarten-age outcomes for many preschool-aged children with previously-identified language delays or disorders, and raise the interesting question of how to predict which children are likely to resolve their language difficulties. There is some evidence that the likelihood of recovery is greater when challenges are relatively circumscribed and becomes less likely with more broad-based difficulties or when receptive language is implicated (Bishop & Edmundson, 1987; Eadie et al., 2014; LaParo et al., 2004). Eadie et al. (2014), for example, classified children as having a language disorder or typical language at age 4 based on performance on the Clinical Evaluation of Language Fundamentals: Preschool–Second Edition (CELF:P-2; Semel, Wiig, & Secord, 2004), and then at age 5 based on the Clinical Evaluation of Language Fundamentals–Fourth Edition (CELF-4; Semel, Wiig, & Secord, 2003). The results demonstrated considerable movement in diagnostic classification. Of the children that scored below the cutoff (-1.25 *SD*) on both the Receptive and Expressive Indices of the CELF:P-2 (classified as having a mixed impairment), 45% maintained the same status a year later, 23% no longer tested in the impaired range, 21% tested below cutoff on the Expressive Index only, and 11% tested below cutoff on the Receptive Index only. Of those testing below cutoff on a single index at age 4, changes were even more notable. Specifically, of those testing below cutoff at age 4 on the Receptive Index only, 16% maintained that status at age 5, 66% no longer tested in the impaired range, 10% tested in the impaired ranged for both Receptive and Expressive Indices, and 8% tested in the impaired range on the Expressive Index only. Finally, of those testing below the cutoff at age 4 on the Expressive Index only, 23% maintained that status at 5 years, 50% no longer tested in the impaired range, 12% tested in the impaired ranged for both Receptive and Expressive Indices, and 15% tested in the impaired range on the Receptive Index only.

LaParo et al. (2004) examined the persistence of DLD from 3 to 4.5 years of age. At 4 years, children were classified as having DLD if they obtained a standard score of 80 (-1.33 *SD*) or less on the Auditory Comprehension scale, the Expressive Communication scale, or both scales of the Preschool Language Scale (Zimmerman, Steiner, & Pond, 1979). A particularly strong oral language predictor of status at age 4.5 was the receptive language score on the Reynell Developmental Scales (Reynell, 1991) at age 3. In the data that Bishop and Edmundson (1987) reported, a pattern of strong recovery was more clearly evident in the group of children classified as not showing disorders in language comprehension. Furthermore, in the data that Scarborough and Dobrich (1990) reported, the patterns observed for receptive language were much less systematic than those observed for expressive language—that is, there was no clear evidence of plateau and recovery. However, in both the Bishop and Edmundson (1987) and Scarborough and Dobrich (1990) studies, receptive language testing was less comprehensive and consistent than expressive language, limiting the strength of conclusions.

One caveat to the conclusion that improvement in language scores is often seen toward kindergarten age is the observation that the opposite pattern can and does occur: Children may test in the average range at one point in a study and then test below the cutoff for DLD at a later point. In the Eadie et al. (2014) study, while 45% of children classified with DLD (either Receptive Expressive, Receptive-only, or Expressive-only) at age 4 were classified as having typical language at age 5 (n = 59/132), 5% of those classified as having typical language at age 4 were classified with DLD at age 5 (n = 41/813). The difference between these two values becomes far less striking when viewed as raw numbers (59 vs. 41) or proportions out of the total sample of 945: 8% of the total sample scored in the DLD range at age 4 and the typical range at age 5, and 6% showed the reverse pattern. Thus, movement of scores in both directions can be seen. It should be noted, however, that these data were obtained from a population-based sample on the basis of a test score alone. They certainly demonstrate that classifications based on test score cut-points can change across time intervals. For children with clinically-identified DLD, however, the change in assessment outcomes that is most relevant to the current discussion is that in which scores move from the impaired to the typical range at kindergarten age. The available evidence suggests that such a change may occur relatively frequently.

In the context of the illusory recovery hypothesis, findings of normalized test scores raise the question of how confident we can be in evidence of recovery obtained around kindergarten age. The best data to answer this question are data that report later outcomes of children seen as preschoolers and then again closer to

kindergarten age. Aside from the data that Scarborough and Dobrich (1990) and Scarborough (1991) reported for school-age reading outcomes, there are to our knowledge a limited number of longitudinal studies with preschool, kindergarten, and school-age measurement points to address this question.

With respect to late talkers or children identified with expressive language delay at age 2, a recent study examined outcomes at ages 7 and 12 of late talkers whose language skills had moved into the typical range by age 4 (Dale, McMillan, Hayiou-Thomas, & Plomin, 2014). This study focused on a very large cohort—3,598 twin pairs in the United Kingdom—and used a combination of parent report, telephone-based interview, and online testing to judge language status. The children identified with expressive language delay all had reported use of 15 or fewer words on a short-form United Kingdom version of the McArthur Communicative Development Inventory (Fenson et al., 1994) at age 2. Overall, the study did not find evidence for illusory recovery. As a group, children in the lower range of oral language ability at age 4 were at somewhat elevated risk for literacy difficulties in the school years, including children from a control group that did not show early expressive delays. But the children from the early delay group who caught up to their peers maintained their recovered status. Thus, these study results suggest that children classified as late talkers do not appear to exhibit illusory recovery.

Turning to consider preschool-aged children with broader language difficulties, the most comprehensive research program to date followed a clinically-referred sample of children with DLD beginning at age 4, with further measurement points at ages 5.5, 8, and 15 (Bishop & Adams, 1990; Bishop & Edmundson, 1987; Stothard et al., 1998). When the children were ages 4 and 5.5, the researchers obtained measures of phonology, receptive and expressive vocabulary (picture naming and picture pointing), the ability to convey information in short narratives (scoring grammatical completeness and the amount of information provided), MLU, receptive grammar, and a general verbal comprehension scale. Children were classified as having satisfactory language if they obtained no score below the 3rd percentile and no more than a single score below the 10th percentile. At the assessment at age 5.5, 44% of the children classified as having DLD at age 4 were categorized as resolved or having good language outcomes (Bishop & Edmundson, 1987). Language and literacy assessments in Grade 2 indicated that the 5.5-year-old outcomes were by and large maintained: As a group, the children who had shown continued evidence of language disorder at age 5.5 obtained scores that were significantly below those of the control group or normative sample at age 8 on all but one of 11 measures. In contrast, the majority of mean scores at age 8 for the age 5.5 "good outcome" group were not significantly different from the control or normative means. Two scores, receptive syntax and general comprehension, were below that of the control group but still within normal limits (Bishop & Adams, 1990).

When reassessed at 15 years, however, one third of the children who had been classified as resolved at age 5.5 were once again classified with a language disorder (Stothard et al., 1998). Oral language measures at age 15 included word definitions, receptive vocabulary, sentence repetition, picture naming, receptive grammar, and general oral comprehension. The requirements to be considered as having satisfactory language were the same as in the age 4 to 5.5 study: no score below the 3rd percentile and not more than one score below the 10th. At the group level, the average scores of the children with resolved DLD did not differ from those of the control groups on these measures. At the individual level, however, eight of 26 children classified as resolved at age 5;6 failed the established criterion for satisfactory language. Other measures that did not contribute to diagnostic status were also collected, and at the group level, the "resolved" group scored significantly below controls on a reading and spelling composite, phonological awareness, and the ability to repeat novel nonwords.

Thus, the Grade 2 results as interpreted by Bishop and Adams (1990) did not support the idea of recovery as illusory, whereas the age 15 results that Stothard et al. (1998) reported did. There are several challenges, however, that limit interpretation of these results. One puzzle is why renewed challenges were not observed in the Grade 2 study conducted by Bishop and Adams. This may have been because illusory recovery operates across a longer time span than anticipated (Stothard et al., 1998) or, perhaps more likely, because some of the measures in the Grade 2 study failed to capture challenges that were in fact occurring or emerging. On that point, it is important to note that the Grade 2 oral language results in the Bishop and Adams study were only reported at a group level. These results showed significantly lower scores for the resolved group on measures of language comprehension (although not in the impaired range for the group as a whole), but did not indicate the proportion of children in the resolved group with oral language scores within or below the typical range. This is not inconsistent with the group-level results

reported for the age 15 data, while the individual data reported for that age were more revealing of persistent language difficulties.

In addition, some of the measures, such as MLU and language comprehension, that were used to determine language status from ages 4 to 5.5 were affected by plateaus or ceiling effects in the performance of the typically developing comparison children (Bishop & Edmundson, 1987). This challenge is entirely in line with the concerns about plateaus raised by Scarborough and Dobrich (1990). The Grade 2 results reported in the Bishop and Adams (1990) study did not allow an estimation of whether or not these plateaus were still operative. There are credible reasons, then, to suspect that the Grade 2 data or analysis strategy may not have been sufficiently sensitive to reveal reemerging challenges. There were some challenges with the age 15 data as well. For example, at least one measure was not normed for 15-year-olds (i.e., Test for Reception of Grammar, normed up to 12;11; Bishop, 1983) and may have suffered from ceiling effects. Despite these challenges, the assessment approach appeared to be sufficiently sensitive to identify lingering challenges in some of the participants.

Clinical Implications

Given the information obtained from the studies reviewed, what is the best evidence-based course of action to take in the face of typical range language scores at kindergarten age obtained for children with previously-identified language disorders? More research is needed to support firm conclusions. However, we would argue that the accumulated evidence suggests that at the very least, we should be extremely cautious before deciding that a child with a history of language disorder has recovered and is no longer in need of support. Although continued intervention may not be recommended, it may be appropriate to recommend further monitoring. This recommendation is consistent with a recent recommendation by other researchers (e.g., Dale et al., 2014).

Moreover, if assessment results point to age-appropriate abilities, it may be fruitful to consider whether the tools or measures that were used are likely to be sufficiently sensitive to ongoing challenges or whether they are likely to be measuring those areas of ceiling or plateau that may simply mask ongoing needs. Illusory recovery, if it occurs, may well be an artifact of what and how we assess at kindergarten age, rather than a result of the nature of language at kindergarten age in general (Bishop & Edmundson, 1987; Scarborough, 2009; Scarborough & Dobrich, 1990). Indeed, Bishop and Edmundson (1987), in their interpretation of the age 5.5 results, noted that the outcomes may have differed had they used other measures. On this point, Scarborough (2009) argued that, for any given age, the skills most likely to be sensitive to language disorder are those that are in a period of growth or "ascendancy" in the typically developing population. Thus, by considering skills or tasks that continue to show development and change through the preschool to school-age transition, we can identify language measures as more likely to identify ongoing needs. In addition, there may be measures that continue to be sensitive to DLD at preschool age and school age by virtue of their relevance to the characteristic profile displayed by children with language disorders. The current research literature can offer guidance regarding measures that may be sensitive to ongoing or future challenges with oral and/or written language.

Language Sample Measures: Mean Length of Utterance, Simple Versus Complex Sentence Structures, Grammatical Morphemes

Language samples, when appropriately collected, provide one of the most ecologically valid and sensitive measures of children's expressive language abilities. They can be analyzed over time to document progress on a goal, and/or they can be analyzed relative to comparison databases within the Systematic Analysis of Language Transcripts (SALT; Heilmann, Miller, & Nockerts, 2010; Miller, Andriacchi, & Nockerts, 2011) to obtain age-comparison *z* scores on a number of skills. There are many useful measures that can be taken from a language sample, and the research literature provides indications of measures that are likely to either decrease in sensitivity or remain sensitive as children get older.

Mean length of utterance. While MLU is a clinically informative measure, the magnitude of difference in measured MLU between children with typical language and language disorder—and thus its utility as a marker of language strength and challenge—declines over the preschool years (Goffman & Leonard, 2000; Scarborough & Dobrich, 1990). This may be in part because of declining gains in MLU in typical development. Such plateau effects were reported for typically developing children by Bishop and Edmundson (1987) and Scarborough and Dobrich (1990), and also appeared to be in operation in the data that Goffman and Leonard (2000) reported. However, as we will discuss below, MLU collected from a sufficiently rich or challenging context may not be subject to the same concerns. Looking beyond MLU, clinicians can gain useful clinical information by considering not only the length of children's utterances but

the range and sophistication of sentence types that they are using.

Simple versus complex syntax. Language sample data can be examined not only for global indices such as MLU, but also for specific language forms, such as complex sentences. Complex sentences consist of an independent clause plus one or more dependent clauses (Justice & Ezell, 2016). They serve a number of critical functions, such as (but not limited to) adding specificity and discussing relationships of time, conditionality, and causality. Examples of complex sentence forms include those with complement clauses (e.g., *The girl thought that the airplane was moving*), adverbial clauses (e.g., *The girl got angry because the airplane wasn't moving; The airplane finally took off after the runway was cleared*), and relative clauses that add specific information about a noun (e.g., *The airplane that was about to take off taxied down the runway*). In addition, sentences with nonfinite (i.e., non-tensed) verbs may be considered complex, such as infinitives (e.g., *The girl wanted the airplane to take off; the girl wanted the boy to throw the plane*) and gerund clauses (e.g., *The girl got impatient waiting for her turn*).

Complex sentence forms develop over the preschool years and continue to show growth in use over the school years. It is important for clinicians to consider whether they are seeing evidence of complex sentence forms in the speech of children on their caseloads, particularly before deciding that a kindergarten-age child's language development is on track. In typical language development, the early development of complex sentence structures begins soon after children begin to combine words into simple utterances (Arndt & Schuele, 2013). By school age, typically developing children produce many, if not all, complex sentence types facilitating efficient communication about an expanded range of relationships. Thus, they enter school with the ability to combine clauses in complex ways and use this tool increasingly as they get older (see Arndt & Schuele, 2013; Frizelle, Thompson, McDonald, & Bishop, 2018; Nippold, 2007). Researchers have begun to pay attention to the complex sentence abilities of children with DLD. The emergence of complex sentence structures is linked to MLU (e.g., Tyack & Gottsleben, 1986), and thus it is not surprising that this occurs at a somewhat older age in children with DLD (Arndt & Schuele, 2013). In addition to the late emergence, there is evidence that children with DLD may demonstrate a limited range and/or less frequent use of these forms (Tuller, Henry, Sizaret, & Barthez, 2012) and ongoing grammatical errors within complex sentences long after they have disappeared from the speech of typical children (Schuele & Dykes, 2005).

Thus, a child who produces grammatical but simple sentences may very well have ongoing difficulties with language learning and use that will limit his or her academic and social success. Researchers have noted that many clinicians may feel ill-prepared to work with complex syntax (for introductions, see Arndt & Schuele, 2013, or Eisenberg, 2013). In addition, the SALT software package (Miller et al., 2011) makes it possible to compare a language transcript to a typical sample with respect to the complexity of the sentences produced, using a measure called the subordination index. The Edmonton Narrative Norms Instrument (Schneider, Hayward, & Dubé, 2006) provides a scoring system for complex sentences called the Complexity Index as well as a local normative database from narrative samples from children aged 4 to 9.

Grammatical errors. Relatively persistent difficulty with the omission of grammatical morphemes is a hallmark of children with DLD (Leonard, 2014). These include finite verb forms such as past tense *–ed*, third person singular *–s*, and copula and auxiliary forms of be (*is, are, am, was, were*). By kindergarten age, children with typical language development produce these forms at mastery levels, whereas children with language disorders continue to show omissions into the school years (Rice, Wexler, & Hershberger, 1998). Children with DLD often continue to produce grammatical morphemes at rates well below their typically developing peers, as production in typical development reaches a ceiling. Indeed, composite indices of grammatical morpheme errors can reliably differentiate children with DLD from their typically developing peers at ages 5 and 6 (Gladfelter & Leonard, 2013; Guo & Schneider, 2016; Souto, Leonard, & Deevy, 2014). Thus, lingering difficulties with the use of grammatical morphemes at the end of the preschool years, even if they appear to be relatively isolated, may signal ongoing challenges with language learning and use. It is worth noting that the overall percentage of utterances in a sample showing some form of grammatical error has also been found to reliably distinguish children with DLD from their peers into the school years (Guo & Schneider, 2016). For a given elicitation context (e.g., conversation or narration), a comparison of the percentage of utterances within a language sample that contain error, relative to same-age peers, can be obtained using the SALT program (Miller et al., 2011).

Narrative Comprehension and Production

There are a number of good reasons to consider how effectively children are able to work with units of connected text as they are reaching the end of the preschool years. Narrative discourse entails greater cognitive demands than conversation or isolated word and sentence production.

Narrative discourse skills continue to undergo considerable development from preschool to the school years (Schneider et al., 2006), and narrative contexts tend to elicit language that is more representative of emerging or higher-level abilities, such as longer and more syntactically complex sentences or phrases (MacLachlan & Chapman, 1988; Wagner, Nettelbladt, Sahlén, & Nilholm, 2000; Westerveld, Gillon, & Miller, 2004). For these reasons, narrative tasks may be sensitive to ongoing language difficulty, even if simpler or earlier-developing language skills appear to be relatively strong. Indeed, several research studies have shown narrative-based measures to be strong predictors of later language and academic outcomes (Bishop & Edmundson, 1987; Botting et al., 2001; Stothard et al., 1998). In the longitudinal research conducted by Bishop and colleagues (Bishop & Adams, 1990; Bishop & Edmundson, 1987; Stothard et al., 1998), for example, narrative ability at age 4 was the best predictor of oral language outcomes at age 5;6 and reading at age 8. Narrative abilities at age 5;6 also predicted which children would be reclassified from resolved DLD at age 5;6 to impaired at 15 (Stothard et al., 1998).

There are several formal tools available to examine narrative abilities in preschool and/or kindergarten-aged children. The Test of Narrative Language (Gillam & Pearson, 2004) has norms from age 5;0, and the Test of Narrative Language–Second Edition (Gillam & Pearson, 2017) from 4;0[1]. These tests examine both receptive and expressive abilities. The Edmonton Narrative Norms Instrument (Schneider et al., 2006) examines narrative production and provides local norms from age 4;0. Both the Test of Narrative Language and the Edmonton Narrative Norms Instrument have the advantage that the collected narrative samples can be used as language samples to analyze word- and sentence-level aspects of production within SALT (Miller et al., 2011), as there are reference databases for both of these tools. Examining narrative language samples in this way has the potential advantage that plateaus in measures obtained from typically developing children, sometimes seen in conversational or play contexts, may be less likely in narratives. In contrast to conversation, children's stories have been found to exhibit longer sentences, more syntactically complex language, and more phrasal expansions (MacLachlan & Chapman, 1988; Wagner et al., 2000; Westerveld et al., 2004).

Sentence Repetition

Sentence repetition, while quite removed from functional, everyday communication, has been shown to be an excellent marker of language-based difficulties (e.g., Archibald & Joanisse, 2009). Sentence repetition scores obtained in kindergarten have been shown to predict reading outcomes in Grade 2 (Catts, Fey, Zhang, & Tomblin, 2001). Additionally, in a study focused on preschool-aged children with expressive language delays, sentence repetition at age 3 to 4 was the best predictor of whether or not children would continue to show delays when assessed a year later (Everitt, Hannaford, & Conti-Ramsden, 2013). Sentence repetition has also been shown to be a particularly sensitive marker of language disorder in school-age children (Conti-Ramsden, Botting, & Faragher, 2001). The sensitivity of sentence repetition may stem from the fact that it likely draws on a number of abilities that are challenged in language disorder such as memory and facility with language forms (Wiig, Semel, & Secord, 2013). For this reason, difficulty with sentence repetition may be an indicator that a child continues to struggle to work with language efficiently, and thus is likely to struggle as academic and social demands of language use increase, even if he or she shows relative strengths in single word or sentence production, or everyday conversation. Scaled sentence repetition scores are available for several commonly used commercially available tests, including the CELF:P-2 (Semel et al., 2004) and the CELF-5 (Wiig et al., 2013).

Literacy Predictors

Although children heading off to kindergarten may not be reading and writing just yet, they have been building, and continue to build, emergent literacy skills. There are a number of skills that, when assessed at kindergarten age, have been shown to predict later challenges with literacy in children with language disorders. These include rapid automatized naming (repeatedly naming a set of letters, digits, or objects as quickly as possible), phonemic awareness measures, the ability to identify letters of the alphabet (Catts et al., 2001, 2002), and narrative skills (Bishop & Edmundson, 1987; Botting et al., 2001). For these reasons, it may be fruitful for clinicians to closely consider these abilities at kindergarten age in children who continue to struggle with language development, but also in those children who may appear to have resolved their language difficulties.

[1] Clinicians are encouraged to read the manual for the Test of Narrative Language-2 carefully before interpreting obtained scores, as sensitivity of the test to DLD at -1 *SD* is quite low, and the authors recommend a standard score of 92 (-0.5 *SD*) as the optimal threshold for identification.

To summarize, the research literature points to measures that may be particularly useful in the identification of ongoing language needs at kindergarten age or risks of future language and literacy challenges. These include measures of grammatical morphology use, use of complex syntax, narrative production and/or comprehension, phonological awareness, sentence repetition, and rapid naming. Clinicians may want to consider including these measures in kindergarten-age assessments, if not already doing so, particularly for children who have shown gains in their language ability in play/conversation or when tested at the word or sentence level.

It is important to acknowledge that these measures, highlighted as sensitive to language disorder or language/literacy needs have not, by and large, been examined specifically within the context of predicting future outcomes in children whose oral language difficulties seem to have resolved at kindergarten age. More research is needed to examine whether these measures continue to point to language needs or risk, even if other measures (e.g., language measures at the single word or sentence level, or in everyday play or conversation) indicate no concerns.

Limitations

Before concluding, it is important to acknowledge several limitations to the current review and existing literature. First, the classifications that inform research outcomes are typically based on test score cutoffs, a practice that supports consistency and objectivity but also comes with limitations. There may be differences among studies or research and clinical contexts in the cutoff scores that are taken as evidence of impaired or satisfactory language development. The current review described the criteria adopted in some of the more prominent studies. However, it was not intended to report in detail on differences among studies or to adjudicate among different classification schemes. It is clear that different criteria will produce different results with respect to the number of children that appear to have continued language difficulties or appear to have resolved.

Moreover, research using discrete cut-points may not take into account the error inherent in all test scores to qualify or temper conclusions. The research studies cited in this review used discrete cut-points, whereas in clinical practice it is important to consider the confidence interval around a given score (based on the standard error of measurement and given in the test manual) as an indication of the range within which the child's "true score" is likely to lie. Eadie et al. (2014) noted that the potential variability in test scores from ages 4 to 5 (based on the 95% limits to agreement from their dataset, calculated between the CELF:P-2 used at age 4 and the CELF-4 used at age 5) exceeded what would be expected based on the reported standard error of measurement for the tools they used. Future studies in this area would benefit from considering the test-reported standard error of measurement for the scores used for classification at different time points as an indication of confidence in the classifications that were made. Simply put, discrete cut-points may indicate that a child falls within the impaired range at one point and within the average range at another. If the confidence intervals around the scores, however, overlap, then one could be less confident that the scores truly differed.

Clinicians making use of the findings of a particular study may want to consider the likely reliability of the classification system adopted in the study. Did it rely on a single test score for classification or use a number of measures? Did it take into account standard error of measurement for the measures used? Did it adopt a lenient or conservative criterion to identify children as resolved? This paper has presented the argument that more research on kindergarten and later outcomes is needed. For researchers pursing this line of inquiry, conclusions may be strengthened by consideration of outcomes taking into account confidence intervals for the tools used, and/or how outcomes might vary according to the criterion set for a diagnosis of DLD (e.g., -1.25 *SD* vs -2 *SD*; cf. Eadie et al., 2014).

It is also important to keep in mind that the use of test scores in isolation is unlikely to reflect the range of factors that contribute to diagnostic decisions in clinical contexts. Test scores may not reveal functional challenges that also inform diagnoses in clinical contexts (see Charest et al., 2019, for further discussion). The information provided in this review is intended to assist clinicians as they interpret test scores within the context of their functional observations. It may also provide a context to interpret and further investigate any noted discrepancies between test score and functional observations.

Second, one of the main arguments of this review is that assessment outcomes, and the picture of recovery versus continuing disorder, may vary considerably according to the measures used. Clinicians making use of the literature on illusory recovery may want to bear in mind the fact that much of the relevant literature is now several decades old and may have used tools that are no longer current or whose psychometric properties may not have been as extensively developed and documented as more current tools. A larger body of research is needed, using more current tools that have documented acceptable diagnostic accuracy (e.g., see Spaulding, Plante, & Farinella, 2006, regarding sensitivity and specificity).

Finally, this review has not discussed the role of intervention in the trajectory of language disorders and outcomes or how intervention may affect the interpretation of recovery data. When a clinician has collected assessment data from a child that has been receiving intervention, it would be reasonable to assume that the outcomes have been influenced by the treatment. However, while the intervention may have been successful in accelerating learning of treated forms, it may not have resolved the underlying factors contributing to the language disorder (Paul & Norbury, 2012). Thus, normal-range scores on reassessment could reflect at least three realities: (a) true resolution of learning challenges, (b) the illusion of recovery due to the course of typical language development or the nature of the measures used, or (c) treatment-supported gains in learning that may or may not be sustainable once learning supports are removed. The research studies reviewed in this paper were not designed to control the amount or nature of interventions received over the course of the study period. Typical-range scores maintained over some interval within a research study may have been obtained with or without ongoing treatment, and so clinicians and researchers may want to exert extra caution before assuming that a child is likely to maintain gains following the withdrawal of services.

Our research team recently surveyed S-LPs working in Alberta about their practices, perspectives, and questions related to assessment and diagnosis of language disorders (Charest et al., 2019). One of the survey respondents commented: "*Can children with a diagnosed language delay or disorder recover? It may seem simple, but I'm not sure that I even know the answer to this question.*" Following the literature review presented in this paper, we would suggest that the question is not at all simple, but reflects the uncertainty that researchers have acknowledged for a number of years. The challenge with this uncertainty is that clinicians are often faced with interpreting assessment results taken at a specific point in time and using that information to make decisions or predictions regarding children's likely need for supports as they move into new programs or educational settings. The issues raised by the illusory recovery hypothesis (i.e., Scarborough & Dobrich, 1990) bring to mind questions of how lingering (but unidentified) language-based challenges might affect academic and social success in the school years, and whether we can mitigate such effects by continuing to provide language learning or other classroom supports.

The research literature can nonetheless support clinicians in their decisions about assessment and diagnosis at the transition from preschool to school age. The current state of the literature is in line with the view that children with preschool-identified language disorder who test in the average range at kindergarten age may indeed have ongoing language needs. More research is definitely needed. In the absence of clear guidelines, clinicians may want to proceed very cautiously before discharging children from services or monitoring. In children who have made great strides at the word or sentence level or in their communicative success in everyday play and conversation, it may be fruitful to examine skills that are continuing to develop from preschool into the school years in typical development and those that have been shown to be predictive of later language and literacy outcomes.

References

Aram, D. M., Ekelman, B. L., & Nation, J. E. (1984). Preschoolers with language disorders: 10 years later. *Journal of Speech, Language, and Hearing Research, 27,* 232–244. doi:10.1044/jshr.2702.244

Archibald, L. M. D., & Joanisse, M. F. (2009). On the sensitivity and specificity of nonword repetition and sentence recall to language and memory impairments in children. *Journal of Speech, Language, and Hearing Research, 52,* 899–914. doi:10.1044/1092-4388(2009/08-0099)

Arndt, K. B., & Schuele, C. M. (2013). Multiclausal utterances aren't just for big kids: A framework for analysis of complex syntax in spoken language of preschool- and early school-age children. *Topics in Language Disorders, 33,* 125–139. doi:10.1097/TLD.0b013e31828f9ee8

Beitchman, J. H., Brownlie, E. B., Inglis, A., Wild, J., Mathews, R., Schachter, D., … Lancee, W. (1994). Seven-year follow-up of speech/language-impaired and control children: Speech/language stability and outcome. *Journal of the American Academy of Child & Adolescent Psychiatry, 33,* 1322–1330. doi:10.1111/j.1469-7610.1996.tb01493.x

Bishop, D. V. M. (1983). *Test for Reception of Grammar*. Manchester, UK: University of Manchester.

Bishop, D. V. M. (2014). Ten questions about terminology for children with unexplained language problems. *International Journal of Language and Communication Disorders, 49,* 381–415. doi:10.1111/1460-6984.12101

Bishop, D. V. M., & Adams, C. (1990). A prospective study of the relationship between specific language impairment, phonological disorders and reading retardation. *The Journal of Child Psychology and Psychiatry, 31,* 1027–1050. doi:10.1111/j.1469-7610.1990.tb00844.x

Bishop, D. V. M., & Edmundson, A. (1987). Language-impaired 4-year-olds: Distinguishing transient from persistent impairment. *Journal of Speech and Hearing Disorders, 52,* 156–173. doi:10.1044/jshd.5202.156

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & the CATALISE-2 consortium. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *The Journal of Child Psychology and Psychiatry, 58,* 1068–1080. doi:10.1111/jcpp.12721

Botting, N. (2007). Comprehension difficulties in children with specific language impairment and pragmatic language impairment. In K. Cain & J. Oakhill (Eds.), *Children's comprehension problems in oral and written language* (pp. 81–103). New York, NY: Guilford Press.

Botting, N., Faragher, B., Simkin, Z., Knox, E., & Conti-Ramsden, G. (2001). Predicting pathways of specific language impairment: What differentiates good and poor outcome? *The Journal of Child Psychology and Psychiatry, 42,* 1013–1020. doi:10.1111/1469-7610.00799

Catts, H. W., Fey, M. E., Tomblin, B. J., & Zhang, X. (2002). A longitudinal investigation of reading outcomes in children with language impairments. *Journal of Speech, Language, and Hearing Research, 45,* 1142–1157. doi:10.1044/1092-4388(2002/093)

Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (2001). Estimating the risk of future reading difficulties in kindergarten children: A research-based model and its clinical implementation. *Language, Speech, and Hearing Services in Schools, 32,* 38–50. doi:10.1044/0161-1461(2001/004)

Charest, M., Borger, P., Chan, C., Sander, K., Yip, B., McFarlane, L., & Schneider, P. (2019). Assessment, diagnosis and recovery from language disorder at kindergarten age: A survey of clinicians. *Canadian Journal of Speech-Language Pathology and Audiology, 43,* 63–80.

Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *The Journal of Child Psychology and Psychiatry, 42,* 741–748. doi:10.1111/1469-7610.00770

Dale, P. S., McMillan, A. J., Hayiou-Thomas, M. E., & Plomin, R. (2014). Illusory recovery: Are recovered children with early language delay at continuing elevated risk? *American Journal of Speech-Language Pathology, 23,* 437–447. doi:10.1044/2014_AJSLP-13-0116

Eadie, P., Nguyen, C., Carlin, J., Bavin, E., Bretherton, L., & Reilly, S. (2014). Stability of language performance at 4 and 5 years: Measurement and participant variability. *International Journal of Language and Communication Disorders, 49,* 215–227. doi:10.1111/1460-6984.12065

Eisenberg, S. L. (2013). Grammar intervention: Content and procedures for facilitating children's language development. *Topics in Language Disorders, 33,* 165–178. doi:10.1097/TLD.0b013e31828ef28e

Everitt, A., Hannaford, P., & Conti-Ramsden, G. (2013). Markers for persistent specific expressive language delay in 3–4-year-olds. *International Journal of Language and Communication Disorders, 48,* 534–553. doi:10.1111/1460-6984.12028

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59*(5), 1–185. doi:10.2307/1166093

Frizelle, P., Thompson, P. A., McDonald, D., & Bishop, D. V. M. (2018). Growth in syntactic complexity between four years and adulthood: Evidence from a narrative task. *Journal of Child Language, 45,* 1174–1197. doi:10.1017/S0305000918000144

Gillam, R. B., & Pearson, N. A. (2004). *The Test of Narrative Language*. Austin, TX: Pro-Ed.

Gillam, R. B., & Pearson, N. A. (2017). *The Test of Narrative Language* (2nd ed.). Austin, TX: Pro-Ed.

Gladfelter, A., & Leonard, L. B. (2013). Alternative tense and agreement morpheme measures for assessing grammatical deficits during the preschool period. *Journal of Speech, Language, and Hearing Research, 56,* 542–552. doi:10.1044/1092-4388 (2012/12-0100)

Goffman, L., & Leonard, J. (2000). Growth of language skills in preschool children with specific language impairment: Implications for assessment and intervention. *American Journal of Speech-Language Pathology, 9,* 151–161. doi:10.1044/1058-0360.0902.151

Guo, L. Y., & Schneider, P. (2016). Differentiating school-aged children with and without language impairment using tense and grammaticality measures from a narrative task. *Journal of Speech, Language, and Hearing Research, 59,* 317–329. doi:10.1044/2015_JSLHR-L-15-0066

Hammill, D. D., Brown, V. L., Larsen, S. C., & Wiederholt, J. L. (1980). *Test of Adolescent Language*. Austin, TX: Pro-Ed.

Heilmann, J. J., Miller, J. F., & Nockerts, A. (2010). Using language sample databases. *Language, Speech, and Hearing Services in Schools, 41,* 84–95. doi:10.1044/0161-1461(2009/08-0075)

Johnson, C. J., Beitchman, J. H., Young, A., Escobar, M., Atkinson, L., Wilson, B., … Wang, M. (1999). Fourteen-year follow-up of children with and without speech/language impairments: Speech/language stability and outcomes. *Journal of Speech, Language, and Hearing Research, 42,* 744–760. doi:10.1044/jslhr.4203.744

Justice, L. M., & Ezell, H. K. (2016). *The syntax handbook: Everything you learned about syntax…but forgot* (2nd ed.). Austin, TX: Pro-Ed.

LaParo, K. M., Justice, L., Skibbe, L. E., & Pianta, R. C. (2004). Relations among maternal, child, and demographic factors and the persistence of preschool language impairment. *American Journal of Speech-Language Pathology, 13,* 291–303. doi:10.1044/1058-0360(2004/030)

Leonard, L. B. (2014). *Children with specific language impairment* (2nd ed.). Cambridge, MA: MIT Press.

MacLachlan, B. G., & Chapman, R. S. (1988). Communication breakdowns in normal and language learning-disabled children's conversation and narration. *Journal of Speech and Hearing Disorders, 53,* 2–7. doi:10.1044/jshd.5301.02

Miller, J. F., Andriacchi, K., & Nockerts, A. (2011). *Assessing language production using SALT software: A clinician's guide to language sample analysis*. Middleton, WI: SALT Software, LLC.

Nippold, M. A. (2007). *Later language development: School-age children, adolescents, and young adults* (3rd ed.). Austin, TX: Pro-Ed.

Nippold, M. A., & Schwarz, I. E. (2002). Do children recover from specific language impairment? *International Journal of Speech-Language Pathology, 4,* 41–49. doi:10.1080/14417040210001669221

Paul, R. (1996). Clinical implications of the natural history of slow expressive language development. *American Journal of Speech-Language Pathology, 5,* 5–21. doi:10.1044/1058-0360.0502.05

Paul, R., & Norbury, C. F. (2012). *Language disorders from infancy through adolescence: Listening, speaking, reading, writing, and communicating* (4th ed.). St. Louis, MO: Elsevier.

Paul, R., & Roth, F. P. (2011). Characterizing and predicting outcomes of communication delays in infants and toddlers: Implications for clinical practice. *Language, Speech, and Hearing Services in Schools, 42,* 331–340. doi:10.1044/0161-1461(2010/09-0067)

Pence-Turnbull, K.-L., & Justice, L. M. (2012). *Language development: From theory to practice* (2nd ed.). Saddle River, NJ: Pearson Education.

Rescorla, L. (2002). Language and reading outcomes to age 9 in late-talking toddlers. *Journal of Speech, Language, and Hearing Research, 45,* 360–371. doi:10.1044/1092-4388(2002/028)

Rescorla, L. (2005). Age 13 language and reading outcomes in late-talking toddlers. *Journal of Speech, Language, and Hearing Research, 48,* 459–472. doi:10.1044/1092-4388(2005/031)

Reynell, J. (1991). *Reynell Developmental Language Scales* (U.S. ed.). Los Angeles, CA: Western Psychological Services.

Rice, M. L., Wexler, K., & Hershberger, S. (1998). Tense over time: The longitudinal course of tense acquisition in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 41,* 1412–1431. doi:10.1044/jslhr.4106.1412

Scarborough, H. S. (1990). Index of Productive Syntax. *Applied Psycholinguistics, 11,* 1–22. doi:10.1017/S0142716400008262

Scarborough, H. S. (1991). Early syntactic development of dyslexic children. *Annals of Dyslexia, 41,* 207–220. doi:10.1007/BF02648087

Scarborough, H. S. (2009). Connecting early language and literacy to later reading (dis)abilities: Evidence, theory and practice. In F. Fletcher-Campbell, J. M. Soler, & G. Reid (Eds.), *Approaching difficulties in literacy development: Assessment, pedagogy and programmes* (pp. 23–38). London, UK: Sage.

Scarborough, H. S., & Dobrich, W. (1990). Development of children with early language delay. *Journal of Speech, Language, and Hearing Research, 33,* 70–83. doi:10.1044/jshr.3301.70

Schneider, P., Hayward, D., & Dubé, R. V. (2006). Storytelling from pictures using the Edmonton Narrative Norms Instrument. *Journal of Speech-Language Pathology and Audiology, 30,* 224–238.

Schuele, C. M., & Dykes, J. C. (2005). Complex syntax acquisition: A longitudinal case study of a child with specific language impairment. *Clinical Linguistics & Phonetics, 19,* 295–318. doi:10.1080/02699200410001703709

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical Evaluation of Language Fundamentals* (4th ed.). San Antonio, TX: Pearson.

Semel, E., Wiig, E. H., & Secord, W. A. (2004). *Clinical Evaluation of Language Fundamentals–Preschool* (2nd ed.). San Antonio, TX: Pearson.

Souto, S. M., Leonard, L. B., & Deevy, P. (2014). Identifying risk for specific language impairment with narrow and global measures of grammar. *Clinical Linguistics & Phonetics, 28,* 741–756. doi:10.3109/02699206.2014.893372

Spaulding, T. J., Plante, E., & Farinella, K. A. (2006). Eligibility criteria for language impairment: Is the low end of normal always appropriate? *Language, Speech, and Hearing Services in Schools, 37,* 61–72. doi:10.1044/0161-1461(2006/007)

Stark, R. E., Bernstein, L. E., Condino, R., Bender, M., Tallal, P., & Catts, H. (1984). Four-year follow-up study of language impaired children. *Annals of Dyslexia, 34,* 49–68. doi:10.1007/BF02663613

Stothard, S. E., Snowling, M. J., Bishop, D. V. M., Chipchase, B. B., & Kaplan, C. A. (1998). Language-impaired preschoolers: A follow-up into adolescence. *Journal of Speech, Language, and Hearing Research, 41,* 407–418. doi:10.1044/jslhr.4102.407

Tomblin, J. B., Zhang, X., Buckwalter, P., & O'Brien, M. (2003). The stability of primary language disorder: Four years after kindergarten diagnosis. *Journal of Speech, Language, and Hearing Research, 46,* 1283–1296. doi:10.1044/1092-4388(2003/100)

Tuller, L., Henry, C., Sizaret, E., & Barthez, M.-A. (2012). Specific language impairment at adolescence: Avoiding complexity. *Applied Psycholinguistics, 33,* 161–184. doi:10.1017/S0142716411000312

Tyack, D. L., & Gottsleben, R. H. (1986). Acquisition of complex sentences. *Language, Speech, and Hearing Services in Schools, 17,* 160–174. doi:10.1044/0161-1461.1703.160

Wagner, C. R., Nettelbladt, U., Sahlén, B., & Nilholm, C. (2000). Conversation versus narration in pre-school children with language impairment. *International Journal of Language & Communication Disorders, 35,* 83–93. doi:10.1080/136828200247269

Westerveld, M. F., Gillon, G. T., & Miller, J. F. (2004). Spoken language samples of New Zealand children in conversation and narration. *Advances in Speech-Language Pathology, 6,* 195–208. doi:10.1080/14417040400010140

Whitehurst, G. J., & Fischel, J. E. (1994). Practitioner review: Early developmental language delay: What, if anything, should the clinician do about it? *The Journal of Child Psychology and Psychiatry, 35,* 613–648. doi:10.1111/j.1469-7610.1994.tb01210.x

Wiig, E. H., Semel, E., & Secord, W. A. (2013). *Clinical Evaluation of Language Fundamentals* (5th ed.). Bloomington, MN: Pearson.

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (1979). *Preschool Language Scale* (Revised Edition). San Antonio, TX: The Psychological Corporation.

Authors' Notes

Correspondence concerning this article should be addressed to Monique Charest, Department of Communication Sciences and Disorders, 2-70 Corbett Hall, University of Alberta, Edmonton, AB, Canada, T6G 2G4. Email: mcharest@ualberta.ca

Acknowledgments

No funding was received for this work.

Disclosures

No conflicts of interest, financial or otherwise, are declared by the authors.

Assessment, Diagnosis, and Recovery from Language Disorder at Kindergarten Age: A Survey of Clinicians

Évaluation, diagnostic et récupération du trouble du langage chez les enfants en âge de commencer la maternelle : un sondage auprès de cliniciens

KEYWORDS

LANGUAGE DEVELOPMENT
LANGUAGE ASSESSMENT
DIAGNOSIS
LANGUAGE DELAY
LANGUAGE IMPAIRMENT
LANGUAGE DISORDER
PRESCHOOL
ILLUSORY RECOVERY
DEVELOPMENTAL LANGUAGE DISORDER
SPECIFIC LANGUAGE IMPAIRMENT

Monique Charest
Paige Borger
Carolyne Chan
Kaitlin Sanders
Beatrice Yip
Lu-Anne McFarlane
Phyllis Schneider

Abstract

There are unanswered questions in the research literature about the long-term trajectories of language disorders that are diagnosed in the preschool years but seem to resolve around kindergarten age. There is some evidence that children tend to maintain their gains in language abilities. On the other hand, there is also some evidence that language difficulties may resurface, suggesting that the initial recovery was *illusory* (Scarborough & Dobrich, 1990). In order to provide clinical context for future research on this topic, we conducted a survey of clinicians in Alberta about their practices and perspectives with regard to diagnosis of language disorders and predictions of future needs. The results revealed perspectives and experiences in line with and inconsistent with concerns about illusory recovery and highlighted challenges with diagnosis and prediction of outcomes within current service delivery contexts. Finally, the results highlighted differences between clinical diagnostic practices and the approach typically taken in research studies. Implications of these differences for interpretation of the research literature and research planning are discussed.

Monique Charest, Paige Borger, Carolyne Chan, Kaitlin Sanders, Beatrice Yip, Lu-Anne McFarlane and Phyllis Schneider

University of Alberta, Edmonton, AB, CANADA

Abrégé

Plusieurs questions concernant la trajectoire à long-terme des enfants diagnostiqués avec un trouble du langage pendant la période préscolaire mais dont le trouble semble se résorber lorsqu'ils sont en âge d'entrer à la maternelle demeurent sans réponse dans la littérature. Certaines données suggèrent que ces enfants auraient tendance à maintenir leurs acquis au plan langagier. D'autres données indiquent quant à elles que les difficultés langagières pourraient réapparaître, suggérant ainsi que leur récupération initiale ne pourrait qu'être illusoire (Scarborough et Dobrich, 1990). Afin de fournir un contexte clinique pour les futures recherches sur le sujet, nous avons effectué un sondage auprès de cliniciens travaillant en Alberta à propos de leurs pratiques cliniques et de leurs perspectives quant au diagnostic de trouble du langage et quant à l'identification des besoins futurs des enfants. Les résultats montrent que certaines perspectives et expériences des cliniciens supportent le phénomène de récupération illusoire, alors que d'autres le réfutent. Cela met en lumière les défis liés au diagnostic du trouble de langage et à l'identification de leurs besoins futurs dans le contexte actuel de prestation de services. Enfin, les résultats mettent en lumière les différences entre ce qui est généralement effectué en clinique et en recherche pour diagnostiquer le trouble de langage. Les implications découlant de ces différences, tant pour l'interprétation de la littérature que pour la planification de futures recherches, sont discutées.

Several decades ago, Scarborough and Dobrich (1990) introduced the concept of *illusory recovery* to account for a puzzling inconsistency in the research literature. Existing studies of children with early-identified developmental language delays/disorders (hereafter referred to as developmental language disorders, or DLD) reported relatively high rates of recovery or normalization in language scores around kindergarten age, whereas other studies, examining outcomes in the school years and beyond, reported longer lasting difficulties. Scarborough and Dobrich suggested that some of the recovery measured around kindergarten age may have reflected the *illusion* of recovery only or a temporary normalization of language scores without resolution of the underlying learning difficulty. They noted that this illusion could result from periods of plateau in typical language growth that allow children on a slower trajectory of learning to temporarily catch up, only to be left behind as language demands and typical development once again accelerate. They also noted that such an illusion could be more or less likely to occur as a function of the developmental sensitivity of the measures used around kindergarten age (see also Bishop & Edmundson, 1987, and Scarborough, 2009). Importantly, the expectation derived from this hypothesis is that given the continued underlying learning difficulties, significant language and/or literacy challenges are likely to re-emerge during the school years (Scarborough & Dobrich, 1990), thus explaining the greater apparent persistence of DLD when measured in the school years.

The issues raised by a possible illusory recovery phenomenon have both theoretical and clinical implications. They are relevant to our understanding of the nature of language growth and language disorders, and the nature of the changes that occur as a function of intervention. From a clinical perspective, the idea of illusory recovery presents a potential interpretive dilemma for clinicians, namely how to proceed given re-assessment results indicating that a child with a previously identified language disorder has achieved typical-range language abilities. On the one hand, the child's achievements should rightly be celebrated. On the other hand, given the assessment results, a clinician may wonder whether it is appropriate to close down a file or pursue some other course of action, such as further review and monitoring. Are the gains in language development likely to be maintained, even as the child is faced with the growing language and literacy demands of the academic environment?

Charest et al. (2019) recently conducted a review of the literature regarding kindergarten-age resolution of DLD (around age 5), with a focus on evidence either indicating that such recovery (if it occurs) tends to be maintained or that language and/or literacy difficulties do in fact tend to resurface. The review pointed to a somewhat limited and equivocal evidence base. Two longitudinal British studies offered particularly relevant findings, as they followed clinically identified children who were assessed as preschoolers, and then again at kindergarten age and school age. In one study, 4-year-old children were selected for inclusion because a speech-language pathologist (S-LP) or pediatrician had identified them as having DLD (Bishop & Edmundson, 1987). They were then seen again at ages 5.5, 8 (Bishop & Adams, 1990), and 15 (Stothard, Snowling, Bishop, Chipchase, & Kaplan, 1998). In another study, 3.5-year-old children were selected because their parents had concerns about their language development or they were deemed at risk for literacy difficulties due to a family history of dyslexia (Snowling, Duff, Nash, & Hulme, 2016). These children were seen again at ages 5.5 and 8.5. Each study adopted its own criteria for a diagnosis of language disorder, and they differed regarding whether or not they additionally classified participants according to non-verbal cognitive scores.

In the cohort of children studied by Bishop and colleagues (Bishop & Adams, 1990; Bishop & Edmundson, 1987; Stothard et al., 1998), kindergarten-age outcomes were relatively good for those children with non-verbal cognition scores in the average range: 44% were classified as having resolved their language disorder at age 5.5. Outcomes at age 8, reported at the group level only, showed that by and large the means of the "resolved" group for language and literacy measures were in the average range, with exceptions for a few measures (Bishop & Adams, 1990). The overall performance of the age 5 resolved group was strong enough at age 8 for the authors to conclude that the age 5 gains had been maintained. At the age 15 assessment, the researchers once again undertook a process of classification at the individual level and concluded that there was evidence for re-emergence of DLD in roughly one third of children classified as recovered at kindergarten age. Thus, there was some evidence for illusory recovery, albeit across a potentially longer time frame than anticipated.

In the cohort described by Snowling et al. (2016), 22% of children identified as having a language disorder at age 3.5 were classified as resolved at age 8. At age 5.5, group means on the language measures revealed a trend toward the eventual age 8 outcomes: On the whole, the group eventually classified as resolved at age 8.5 obtained mean age 5.5 language scores below those of the typically developing group but above those of the group whose DLD persisted at age 8.5. However, the age 5.5 data were

reported at the group level only. Thus, it is not possible to know what proportion of individual children in the age 8.5 recovered and persisting groups would also have been classified as recovered or persisting at kindergarten age.

On the encouraging side, the results from both cohorts (Bishop & Adams, 1990; Bishop & Edmundson, 1987; Snowling et al., 2016) pointed to consistency in language gains from ages 5 to 8 in children with previous diagnoses of DLD. Moreover, Snowling et al. (2016) reported that having language skills in the average range at the kindergarten-age assessment was a positive indicator for developing reading skills commensurate with age expectations in Grade 2. Both studies, however, also reported sub-clinical language and literacy weaknesses at Grade 2 in the recovered group, meaning that group means on some measures were within the typical range, but nonetheless still below the mean scores of typically developing control groups. And, while Stothard et al. (1998) reported evidence for re-emergence of DLD at age 15, Snowling et al. noted that renewed challenges may be yet to be seen in their participants.

The Current Study

There are a number of clinically-relevant questions that merit further research. These include further investigation of outcomes of DLD around kindergarten age and identification of measures that might be most sensitive to ongoing language difficulties. They also include investigation of the potential time course over which language and literacy difficulties may re-emerge over the school years—if at all—as well as further consideration of the potential impacts of sub-clinical weaknesses on academic and social functioning. Much, but not all, of the literature addressing the idea of illusory recovery is several decades old (see Charest et al., 2019, for further discussion). The purpose of the current study was to provide context for future research on these topics and initiate a clinical discussion regarding assessment practices and intervention recommendations at kindergarten age. We sought to obtain a snapshot of current clinical practices and perspectives with respect to assessment around kindergarten age and issues relevant to the illusory recovery hypothesis. The goal was to identify points of alignment and difference between the approaches and evidence that are reported in the research literature and the approaches and perspectives emerging from clinical practice. In particular, we sought information about assessment practices and how decisions regarding diagnosis and recovery are made, as well as whether or not clinicians report experiences or concerns that are relevant to the illusory recovery question.

Method

Participants

Eligible participants were S-LP registrants with the Alberta College of Speech-Language Pathologists and Audiologists who self-identified as working with children. There were 46 respondents. Further information about the respondents is reported under the *Information on Work Context* heading in the Results section.

Materials

The materials consisted of a 13-question online survey. Survey questions included Likert scales, open-ended and limited-choice questions, and combinations of these. The complete survey is presented in the Appendix. Questions 2–4 collected broad information about the respondents' work contexts. Questions 5–9 collected information about practices and opinions with respect to assessment, diagnosis, and service delivery: the types of information that are prioritized when making a diagnosis (Question 5); the score cutoffs that are considered and the tests most commonly used, if tests are used (Questions 6 and 7); confidence in tools to diagnose language delays/disorders and predict future language/literacy needs at 4–6 years (Questions 8a and 8d); the course of action when information sources provide conflicting information (Question 8b); how recovery is identified (Question 8c); and what, if any, age ranges pose a greater challenge for determining whether or not recovery has occurred (Question 9). Questions 10a and 10b asked about clinicians' perceptions of the re-emergence of language and literacy challenges at school age. Finally, questions 11–13 invited clinicians to share any further observations and questions related to assessment, diagnosis, and prediction of risk when considering language ability in the 4- to 6-year-old age range.

Procedures

The research methods were approved and conducted in accordance with the requirements of the Research Ethics Board at the University of Alberta (project approval #Pro00045665). An invitation to participate was published in the April, May, and October 2014 issues of the *Communication Matters* newsletter, distributed via email to all registrants of Alberta College of Speech-Language Pathologists and Audiologists. In 2014, there were 894 registrants who reported working with pediatric populations (0–16 years; Alberta College of Speech-Language Pathologists and Audiologists, n.d.). Participation was entirely voluntary. No identifying information was collected. The informed consent process included the following description:

The research literature reveals two seemingly contradictory findings about the trajectory of children with developmental language impairments. Some studies suggest that many children achieve normal language by about the age of kindergarten entry. Other studies, however, indicate that most language impairments persist over a much longer time frame. We would like to learn more about the trajectory of developmental language impairments and how we can best predict children's risk of long-term difficulties. As a first step, we would like to learn more about your experiences with diagnosis and prediction of language difficulties.

Results

For a number of questions, respondents could select more than one response option, leading to results that sum to greater than 100%. In addition, respondents could choose to not respond to individual questions as appropriate, which led to some variation in the number of responses to different questions. The number of unique respondents for each question as well as results that sum to more than 100% are indicated in the tables, figures, or text as appropriate.

Information on Work Context

There was broad representation from early childhood through adolescence, with the majority (n = 44, 96%) of respondents reporting working with more than one age category (see **Table 1**). Of the 38 respondents who reported working with kindergarten-age children, 13 (34%) worked with clients from preschool through school age (sometimes up to and beyond junior high). Fifteen (39%) worked with preschool and kindergarten-age children (sometimes including toddlers), but not school age. Six (16%) worked with kindergarten and school-age children (sometimes up to and beyond junior high), but not preschool. Finally, four (11%) respondents reported working with preschool through early elementary ages. Of the eight respondents who did not report working with kindergarten-age children, four reported working with toddlers/preschoolers, three with school-age children, and one with toddlers, preschoolers, and school-age children, but not kindergarten age.

Table 1

Question 2: Age of Children on Respondents' Caseloads

Category	Respondents *n* (%)
Toddler (1–2 years old)	18 (39)
Preschool age (3–4 years old)	37 (80)
Kindergarten age (5–6 years old)	38 (83)
Early elementary (7–9 years old)	27 (59)
Late elementary (9–12 years old)	22 (48)
Junior high school and beyond (12 years +)	17 (37)
N unique respondents	46

Note. Respondents could select more than one response category. Responses sum to more than 100%.

Respondents revealed a broad range of work settings. Question 3 was presented as an open response, and we were able to capture the variety of responses with 26 (57%) working in schools (including community health contracts within schools), 13 (28%) in preschools and Early Learning Centres, 13 (28%) in private practice and contract, 11 (24%) in community health (not in schools), and 5 (11%) in hospitals. The results sum to more than 100% as many respondents reported several work settings. The majority of respondents (n = 39, 85%) reported working with children across the range of severity.

Directed Questions

What kinds of information do you consider when making a decision about a child's diagnosis and need for intervention? In Question 5, respondents ranked the seven options, with 1 indicating the most heavily considered piece of information. Categories that were left blank were given a rank of 8. Respondents were able to assign the same ranking to more than one category. **Table 2** presents the average rankings for each of the response options; rankings closer to 1 indicate relatively greater priority. As can be seen, clinicians reported relying most heavily on their clinical observations, followed by standardized tests and parent concerns.

Table 2

Question 5: Prioritization of Information Considered when Making Diagnostic Decisions and Recommendations for Intervention

Information Source	Mean Rank (*SD*)	% Ranked as Most Important
Clinical observations in context	2.1 (1.5)	52
Standardized tests	2.9 (1.7)	24
Parent concerns	3.1 (1.9)	26
Teacher concerns	4.0 (2.1)	9
Language sample analysis	4.9 (2.0)	7
Concerns from other team members	5.2 (2.1)	4
Criterion referenced tests	5.4 (2.2)	4
Other	7.7 (1.1)	0
N unique respondents	46	

Note. Information sources were ranked from 1–8, with 1 indicating the information source that is most heavily considered. Mean ranks closer to 1 indicate relatively greater importance. Respondents could indicate tied rankings if desired. Several respondents ranked more than one option as 'top' priority, and therefore the % responses sum to more than 100%.

If you use standardized tests, what cut-off criteria do you use (in standard deviation and/or percentile rank) for the diagnosis of a language delay/disorder? Responses to Question 6 revealed substantial consistency among respondents, with 33 (72%) indicating the 16th percentile or one standard deviation below the mean. An additional six (13%) reported following the guidelines set in the test manual, and six reported other cutoff scores. Thirteen of the respondents also indicated that their interpretation of test scores will depend on the integration of test results with other information, such as clinical observations and parent report.

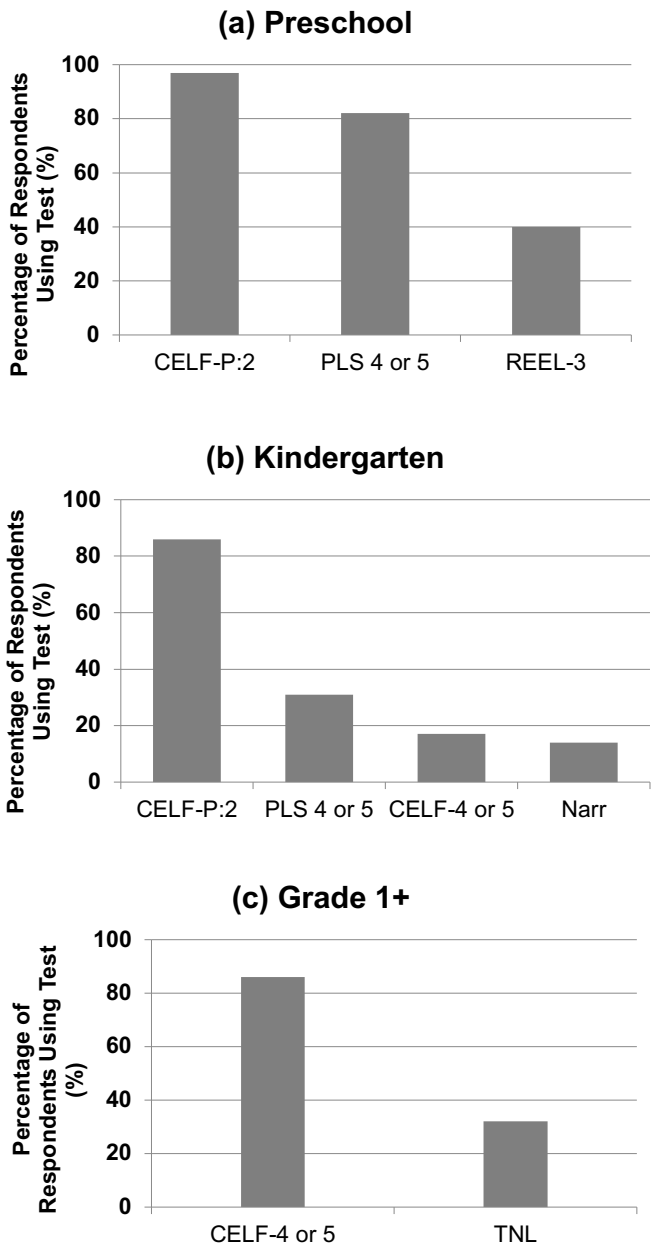
If you use standardized or criterion-referenced tests, please list the tests that you rely on most frequently. In Question 7, respondents listed up to five tests, in decreasing order of frequency of use, for each of three age groups. Although more than 25 different tools were listed across the three age ranges (including some that were not tests), the responses also indicated a consistent core group of tests used at each age. There were five tools that were listed by 30% or more of respondents for at least one age range: the Receptive-Expressive Emergent Language Test–3 (REEL-3; Bzoch, League, & Brown, 2003); the Clinical Evaluation of Language Fundamentals–Preschool:2 (CELF-P:2; Semel, Wiig, & Secord, 2004), The Clinical Evaluation of Language Fundamentals–4 or 5 (CELF-4/5; Semel, Wiig, & Secord, 2003; Wiig, Semel, & Secord, 2013), the Preschool Language Scale–4 or 5 (PLS-4/5; Zimmerman, Steiner, & Pond, 2002, 2011), and the Test of Narrative Language (TNL; Gillam & Pearson, 2004).

Figure 1 presents the distribution of responses for preschool, kindergarten, and Grade 1 and older, respectively, for these five tools. As can be seen, the CELF-P:2 and CELF-4/5 are the most heavily used tests, and many clinicians use the PLS-4/5 in the preschool years. Beyond that, the responses suggest greater similarity in the tools selected for preschool and kindergarten than kindergarten and Grade 1. The CELF-P:2 is selected more often than the CELF-4/5 at kindergarten age, even though both are appropriate to the age range. Narrative tests and the CELF-4/5 begin to receive mention at kindergarten age, although all mentions of narrative, including the TNL, the Edmonton Narrative Norms Instrument (Schneider, Dubé, & Hayward, 2005), and The Renfrew Bus Story (Cowley & Glasgow, 1994), combined together, summed to only approximately 13% of respondents. The shift to greater use of these measures occurs at school age.

Questions 8a, 8b, 8c, and 8d asked specifically about practices when working with children in the 4- to 6-year-old age range.

How confident are you in the tools that you have at your disposal to accurately identify whether or not a child has a language delay/disorder? **Figure 2** presents the distribution of responses to Question 8a, ranging from 1 (*not at all confident*) to 5 (*very confident*). The majority of respondents expressed confidence in their ability to identify language disorders, with a modal rank of 4 (19 of 43 respondents, 44%). Additionally, 13 respondents (30%) selected the highest rating of 5.

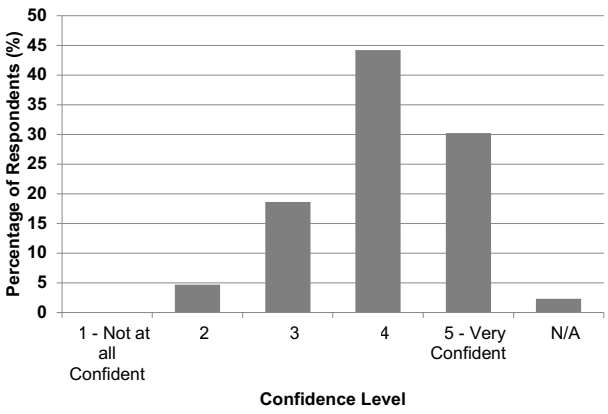
Figure 1



Question 7: Tests most frequently used to assess language in (a) preschool-, (b) kindergarten-, and (c) school-age children. Each respondent could list up to five tests per age group. If the edition of a test was not specified, the most recent version was assumed. CELF-P:2 = Clinical Evaluation of Language Fundamentals Preschool – Second Edition (Semel, Wiig, & Secord, 2004), PLS-4 = Preschool Language Scale – Fourth Edition (Zimmerman, Steiner, & Pond, 2002), PLS-5 = Preschool Language Scale – Fifth Edition (Zimmerman, Steiner, & Pond, 2011), REEL-3 = Receptive-Expressive Emergent Language Test – Third Edition (Bzoch, League, & Brown, 2003), CELF-4 = Clinical Evaluation of Language Fundamentals – Fourth Edition (Semel, Wiig, & Secord, 2003), CELF-5 = Clinical Evaluation of Language Fundamentals – Fifth Edition (Wiig, Semel, & Secord, 2013), TNL = Test of Narrative Language (Gillam & Pearson, 2004). "Narr" (Narrative) = TNL, Edmonton Narrative Norms Instrument (Schneider et al., 2005), or The Renfrew Bus Story (Cowley & Glasgow, 1994).

How do you proceed if there is a discrepancy between the information provided by formal tools (standardized and/or criterion referenced tests) and your clinical judgment about diagnosis and need for intervention? **Table 3** presents the distribution of responses to Question 8b. The modal response (25 of 42 responses, 60%) indicates that clinicians would recommend intervention if either test scores or their observations suggested the need. An additional 13 respondents (31%) indicated that they would only recommend intervention if warranted by clinical observations. None of the respondents indicated that they would only recommend intervention if test scores warranted.

Figure 2



Question 8a: Confidence in tools to identify language delay/disorder, 4–6 year-olds. *n* = 43 respondents.

How do you decide whether a child has recovered from an earlier diagnosis of language delay/disorder? Potential responses to Question 8c were not pre-determined, and so predictably, there was considerable variety in the responses. Most responses mentioned more than one information source. However, most of the responses could be categorized according to the relative emphasis: Emphasis on requiring or integrating positive evidence from formal and informal information sources (16 of 42 responses, 38%), emphasis on parent and/or teacher report of concerns (7 responses, 17%), emphasis on observations of functioning (5 responses, 12%), emphasis on results of re-testing (4 responses, 10%), emphasis on treatment progress (3 responses, 7%), and other (7 responses, 17%). For several of the responses coded within the category of *other*, respondents noted that they either rarely see children who have recovered or are reluctant to assign such a label. One respondent noted, for example, "I wouldn't consider them 'recovered,' just improved...."

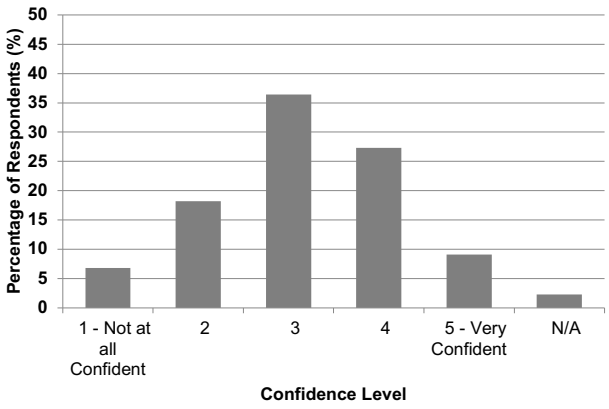
Table 3

Question 8b: Course of Action when Test Scores and Clinical Judgment Provide Discrepant Information About Diagnosis and Need for Intervention

	Respondents *n* (%)
Recommend (or continue) intervention if either information source warrants	25 (60)
Recommend (or continue) intervention if test results warrant	0 (0)
Recommend (or continue) intervention if clinical observations warrant	13 (31)
Conduct further assessment	1 (2)
Discharge client/do not recommend intervention, but share concerns with parent	0 (0)
Discharge client/do not recommend intervention, no further action	0 (0)
Other	3 (7)
N/A – I do not work with children in this age range	1
N unique respondents	43

How confident are you in the tools that you have at your disposal to predict a child's risk for future communication difficulties (i.e., during the school-age years)? **Figure 3** presents the distribution of responses to Question 8d, again from 1 (*not at all confident*) to 5 (*very confident*). The modal rank was 3 (16 of 44 responses, 36%), suggesting a generally neutral assessment of confidence. Twelve respondents (27%) selected a rank of 4, indicating confidence. In contrast to Question 8a, only 4 respondents (9%) chose the highest rating of 5, whereas 11 (25%) chose a confidence ranking of 2 or lower.

Figure 3



Question 8d: Confidence in tools to predict a child's risk of future communication difficulties in school years. *n* = 44 respondents.

Are there any age ranges that you find more difficult than others to evaluate when determining whether recovery from a language delay/disorder has been achieved? If so, which ones? **Table 4** presents the responses to Question 9, both as the raw number of respondents who selected each option, and also expressed as a percentage of individuals who reported working with children in each age range in Question 2. The modal response was *N/A*. Almost all (*n* = 14) of the 15 respondents who selected *N/A* reported working with more than one age group. This suggests that many S-LPs work with multiple age groups, but do not perceive any particular age as being more difficult than others with respect to the determination of recovery. When taken as a percentage of respondents working with different age groups, the age group that appears to pose the greatest difficulty is toddlers. Questions 10a and 10b asked specifically about practices when working with children in the school age range.

If you work with school-age children, how often do you receive referrals for children who had previously been discharged from S-LP services? **Table 5** presents the distribution of responses to Question 10a. The majority of respondents chose *sometimes* (modal response, 17 of 31 responses; 55%) or *often* (10 responses; 32%).

Table 4

Question 9: Age Ranges Posing Greatest Challenge to Determination of Recovery

	Respondents *n*	As % who Work within Age Range
Toddler (1–2 years old)	10	56
Preschool age (3–4 years old)	6	16
Kindergarten age (5–6 years old)	11	29
Early elementary (7–9 years old)	8	30
Late elementary (9–12 years old)	7	32
Junior high school and beyond (12 years +)	6	35
N/A	15	
N unique respondents	46	

Table 5

Question 10a: Perceived Frequency of Re-Referral at School Age

	Respondents *n* (%)
Never	1 (3)
Rarely	1 (3)
Sometimes	17 (55)
Often	10 (32)
Very often	0 (0)
I don't know/am unsure	2 (6)
N/A – I do not work with this age range	12
N unique respondents	43[a]

Note. [a]There were 3 respondents who did not respond to this question because none of them worked with school-age children.

What do you feel are the principal concern(s) that prompt referrals to S-LP services for school-age children who had previously been discharged from S-LP services? For Question 10b, S-LPs ranked their top three perceived concerns, with 1 indicating the concern that they consider to be most frequent. For each respondent, categories that were not selected were given a rank of 6. A small number of respondents ranked all of the response options (from 1–6); these rankings were included in the calculations of means. **Table 6** presents the mean rank assigned to each response category. When we consider the mean rank given to the different response options as well as the number of respondents who selected a given option as the concern of greatest frequency, concerns with literacy are prominent, followed by receptive language and speech intelligibility.

Thematic Analyses

Questions 11–13 were designed as open-ended questions to further probe clinicians' perceptions and practices surrounding the preschool to school-age transition. Responses to these questions were analyzed for thematic content (Braun & Clarke, 2006). Four of the authors read through and familiarized themselves with the responses, developed initial codes to flag pertinent or prevalent ideas, grouped these initial codes into themes, and then revisited the raw data to map how well the themes captured the data, revising themes accordingly. The resulting thematic map reveals three interrelated themes: Knowledge-Support, Challenges, and Concerns, each of which has several sub-themes.

In the first theme, Knowledge-Support, clinicians identified needs for research-based information that would support greater diagnostic and predictive confidence. Although Question 11 asked directly about knowledge needs, many of the comments grouped under this theme were not simply direct replies to this question, but were offered in response to the more open-ended questions 12 and 13. The second theme, Challenges, captures comments regarding system-based factors that impede diagnostic and predictive confidence. Finally, the third theme, Concerns, captures concerns raised about the assessment and identification process, particularly in relation to school-age children. Each of these themes and their sub-themes are described in turn.

Knowledge-support. Clinicians' comments on informational knowledge needs fell into two sub-themes: (a) better understanding of the trajectory of language disorders and the oral language skills or measures that can best predict outcomes and (b) better understanding of how oral language skills affect, and are affected by, children's development across domains such as academics, literacy, and cognitive and social development. The following comments are reflective of this theme[1]:

[1] (Regarding resource/knowledge needs) "Better indicators of risk factors for future difficulties with higher level language and literacy difficulties when assessing toddlers and preschool children."

[2] "Perhaps we need better benchmarks to identify what characteristics are more likely to show a persistent problem with communication versus recovery. I am not sure how you could do this but perhaps there was something like Red Flags for consistent language delays?"

Table 6

Question 10b: Perceived Reasons for Re-Referral at School Age

	Mean Rank *(SD)*	*n* Ranked as Most Frequent[a]
Reading and written language	2.8 (1.9)	10
Listening/ comprehension	3.1 (1.9)	7
Speech intelligibility	3.5 (2.2)	8
Academic achievement	4.1 (1.8)	3
Peer interactions	4.4 (1.7)	2
Other	5.6 (1.4)	2
N unique respondents	30	

Note. Respondents were asked to select up to three concerns, with 1 indicating the perceived most frequent concern. A small number of respondents ranked all response options. Unranked options were given a rank of 6. Mean ranks closer to 1 indicate relatively greater importance.

[a]One respondent indicated a 3-way tie for most frequent area of concern. The values in this column therefore sum to 32 rather than 30.

[3] "The knowledge about children with phonological disorders later having difficulty in reading/ phonological awareness skills is available. I think that the development of social language of kids who have been diagnosed with language delays/disorders in preschool is an area we lack information on - and how social language/pragmatics may impact."

Challenges. Comments on challenges to diagnostic and predictive confidence highlighted system constraints and comprised two sub-themes: (a) time and (b) discontinuities in service. In particular, respondents noted that time constraints place limits on the scope and depth of information they can collect, limiting their ability to fully assess all the areas that they view as important, or their understanding of important aspects of the child's unique experience and environment. Respondents also noted that discontinuity in service provision is an impediment to predicting future outcomes. This theme is reflected in the following comments:

[4] "Would be beneficial to have more studies looking at outcomes and our abilities to predict. We see children at preschool age and then don't have long term info to know how they have done, no way to improve on predictions if we don't have long term effect info."

[5] "We often lack the specific information from the parents and teacher about the whole picture of the child because we only have time to really deal with the immediate matters at hand. Sometimes we do not have all the important information from a parent about the child's history, medical, or even developmental milestones."

[6] "I think that this can be quite a tricky thing to do. While I make predictions, I don't get to follow my clients through elementary school, so I don't often find out whether my predictions are accurate...."

[7] "It is difficult to predict the future risk of children with a history of language delays/disorders when the typical level or type of service delivery to the child changes so dramatically when they enter Grade 1."

Concerns. The final theme that emerged in the respondents' comments can best be characterized as concerns related to the process of assessment and identification, and contains sub-themes related to test concerns and child concerns. With regard to test concerns, respondents expressed mixed feelings about tests. A strong theme was worry that tests do not always reflect or align with children's functional communication challenges and concern that tests may underestimate some children's communication difficulties. At the same time, some respondents emphasized that tests can provide an important source of information to complement observations. Also captured within this

[1] We have corrected typos in the responses.

sub-theme was the view that the profession would benefit from better tools overall. The following quotes are illustrative of this sub-theme:

> **[8]** "Time is limited, and it is easy to fall to the easy way out by using standardized testing to say a child has recovered from and/or has language within average ability at the 16th percentile, even when our observations tell us that a child is not communicating effectively…."

> **[9]** "I feel experience plays a very big role. But…I feel [standardized testing] is a critical piece of the puzzle. I can't walk into a classroom and determine a child's language impairment solely from observation and teacher or parent report."

With regard to child concerns, respondents expressed concerns about children with language needs "falling through the cracks" in the school years. There were specific concerns about children with less severe language disorders having unidentified present or future difficulties, as well as concern that language difficulties may not be as visible as other concerns. These concerns relate to the risk that needs are going unidentified. A complementary point that emerged within this sub-theme is the concern that identified needs are going unmet. The following quotes are illustrative of this sub-theme:

> **[10]** "Mild/moderate kids are falling through the cracks as they often present as average but as the academic work gets more difficult, they fall further and further behind."

> **[11]** "I think teachers and parents are keen to have their children/students read so they are very aware of phonological awareness difficulties. However the language processing difficulties are not always evident and they may think language concerns have resolved and not refer. I think parents and teachers could be better informed about language processing difficulties that may be evident as kids transition, and social language difficulties."

> **[12]** "I suspect that some of the children who appear to have resolved their language issues within the preschool period are not necessarily identified as having the potential to have language issues that become evident once school aged. It is important that children with a history of language delay are monitored as many go on to have persistent language issues that are more subtle."

> **[13]** "One of my colleagues mentioned recently that she feels that many of our preschool/kindergarten kiddos don't receive further SLP support, but then flag as kids with learning disabilities/reading difficulties, etc. Having spent a bit of time in Grade 1 classrooms, I would tend to believe that this is likely the case. As S-LPs typically aren't in those classrooms, the referrals are based on teacher knowledge of speech/language concerns. In my experience, the kids with articulation concerns were brought up immediately, whereas if I brought up a child who I knew had significant language needs in kindergarten/pre-school, teachers often seemed surprised ('He's okay, he's just disorganized')."

> **[14]** "There are many children getting missed still and being released from group intervention even when they still need intervention because you cannot meet all the children's specific speech or language needs without individualized intervention."

Discussion

The illusory recovery hypothesis raises the possibility that children with DLD may test within normal limits around kindergarten age but then have significant needs re-emerge at a later date because the language learning difficulty had not in fact resolved. The risk for these children is that they will potentially miss out on crucial years of support and that their academic, social, or other challenges will be misunderstood. Although clinical practices and assessment tools may have changed in the decades since this hypothesis was first put forward (Scarborough & Dobrich, 1990), several aspects of the survey data suggest that these concerns remain clinically relevant.

One aspect of the data that aligns with a potential illusory recovery phenomenon is the concern that emerged in the thematic data about children with language needs falling through the cracks in the early school years. While there are a number of system-based reasons why children may fall through the cracks, clinicians also highlighted the fact that language-based challenges may simply be missed or misinterpreted, even for children with histories of DLD, particularly those children who do not present at the severe end of the language continuum or have more visible challenges (e.g., speech production, behaviour). This concern is echoed by researchers who have argued that DLD often risks being *invisible* (Bishop, Snowling, Thompson, Greenhalgh, & CATALISE consortium, 2016), particularly when children are able to provide brief, but socially and pragmatically appropriate responses to questions, and converse simply but grammatically about everyday topics of their choosing or contextualized topics rooted in the here and now (Im-Bolter & Cohen, 2007). And, the majority of respondents who work with school age children reported sometimes or often receiving referrals for children who had previously been discharged from S-LP services.

Although descriptors such as *sometimes* and *often* cannot indicate precisely how often this occurs or why—or indeed if it even occurs to an extent that clinicians would deem unacceptable—the response pattern invites further documentation of how often re-referrals occur, when and why they occur, and whether re-emerging or ongoing needs can be identified from the start of formal schooling.

At least one finding, on the other hand, indicates that clinicians do not perceive concerns that align with the illusory recovery hypothesis. Although we have emphasized the late preschool/kindergarten age as a potential time of illusory recovery risk, the survey responses did not reveal particular concern with this age range. Indeed, in response to the question asking if any age range poses a particular challenge to the determination of recovery, the modal response indicated no single age group, and the next most frequent response (when calculated relative to the number of respondents who work within a given age range) identified challenges identifying recovery in toddlers. For a recent review of outcomes for children with language delays identified at age 2, readers are directed to Paul and Roth (2011). Interestingly, although kindergarten age did not stand out as a uniquely challenging age with respect to identifying recovery, several respondents' spontaneous comments offered insight into the nature of difficulty when it occurs. These comments emphasized the challenge that arises when children can be speaking in full sentences yet still have a language disorder and the interpretive challenge that can arise when a child has yet to be faced with the level of academic and literacy demands that emerge in the later years.

Ultimately, questions of whether or not illusory recovery occurs, how often, and why, will need to be answered by future research. Other information obtained in the survey is relevant to such research. In particular, a message that came through consistently in the survey data is the fact that clinicians do not base diagnostic decisions on test scores alone. Indeed, although clinicians reported using standardized tests, slightly more than half of the respondents indicated that they weigh their contextual observations most heavily in their diagnostic decisions and intervention recommendations. Moreover, when asked about the standard score or percentile cutoffs they use when making diagnostic decisions, approximately one third of respondents provided the requested information but also commented that they consider the test score within the context of other observational and reported information. Respondents also indicated a practice of integrating formal and informal information sources to guide decisions regarding recovery, and the majority of respondents indicated that recommendations for intervention could follow from either test scores or clinical observations.

The emphasis on both formal and informal results has implications for how clinicians interpret the research literature on outcomes, as well as implications for research going forward. In contrast to the reported clinical practice, the research base on recovery has relied heavily on test scores or other objective criteria (e.g., mean length of utterance) as the basis for determining outcome (e.g., Bishop & Adams, 1990; Bishop & Edmundson, 1987; LaParo, Justice, Skibbe, & Pianta, 2004; Scarborough & Dobrich, 1990; Snowling et al., 2016; Stark et al., 1984; Stothard et al., 1998). This approach has a number of advantages within a research study, such as feasibility, reliability, and objectivity, but it means that the decisions reached within these studies could differ in important or systematic ways from the decisions that are typically reached by clinicians. For example, a number of the respondents' comments highlighted a concern that test scores may not always align with functional communication challenges. If the obtained test scores do not reflect functional concerns that are evident within contextual observations, studies may overestimate the likelihood that children will appear to have recovered. Conversely, if a research study employs a formal measure that is sensitive to challenges that are real but may not show up in some contexts (e.g., play or everyday conversation), such a study may be less likely to find evidence of recovery. The mismatch in how diagnostic categorizations are made does not mean the research findings are not clinically informative, but they do need to be interpreted within the context of what they tell us. They can help to guide the interpretation of test scores, and the relative weighting of test scores and other observations, by providing indices of how likely gains in test scores are to be maintained. Future longitudinal research on outcomes would benefit from further consideration of how clinical or functional observations and test scores relate or complement each other.

Clinical Considerations

Several other findings from the survey warrant mention. The first is the observation that the majority of clinicians adopt the 16th percentile or one standard deviation below the mean as a diagnostic cutoff. This cutoff, although arbitrary, is commonly used. In the research reported by Snowling et al. (2016), for example, this cutoff was adopted for the primary outcomes in order to align with clinical practices. Not surprisingly, the authors reported that the apparent frequency of recovery varied considerably when other thresholds were adopted. Although one standard deviation below the mean is commonly used, this threshold may not be the most appropriate for all measures.

Empirically-derived cutoffs that provide the best balance of sensitivity and specificity are available, either in test manuals or research reports, for some tests (see Spaulding, Plante, & Farinella, 2006), including some of those that were reported in this survey as being frequently used. For example, the CELF-P:2 (Semel et al., 2004) reports acceptable sensitivity (.85) and specificity (.82) based on a standard score cutoff of 85 or -1 *SD*. The manual for the CELF-5 (Wiig et al., 2013) also reports strong sensitivity (1.00) and specificity (.91) for a -1 *SD* cutoff. However, the manual additionally reports sensitivity and specificity values for different thresholds, and the authors report that the optimal diagnostic threshold is -1.3 *SD* (standard score 80) as this provides the best balance of sensitivity (.97) and specificity (.97). The manual for the TNL (Gillam & Pearson, 2004) indicates acceptable sensitivity (.92) and specificity (.87) at -1 *SD*. The manual for the recently published Test of Narrative Language–2 (TNL-2; Gillam & Pearson, 2017) reports sensitivity and specificity for different thresholds. The reported diagnostic accuracy for -1 *SD* is not acceptable at .55 for sensitivity (specificity is excellent at .98). The authors report that the optimal threshold for identifying language disorder using the TNL-2 is a standard score of 92 or -0.5 *SD*, associated with sensitivity and specificity of .92. Before adopting -1 *SD* as a diagnostic cutoff, clinicians are encouraged to confirm that this cutoff is appropriate for the test in question.

Although we did not ask about clinicians' use of confidence intervals in the interpretation of test scores, we also note here that confidence intervals (provided in the test manual) offer a crucial piece of interpretative information, acknowledging the error inherent in scores and indicating the range within which the child's "true score" may lie. Clinicians are encouraged to incorporate this information into their reporting and use of test scores, if they are not already doing so.

A second issue warranting mention is the uncomfortable intersection of questions about recovery, referral, and prediction of long-term needs with the frequent reality of service at school age. Two relevant issues emerged from the thematic analysis. First, a number of clinicians reported that time limitations—and the limits to information-gathering that follow—negatively affect their confidence in diagnosis and, to a greater extent, prediction of long-term outcomes. On this point, the test-focused approach taken in research studies may be particularly relevant if it can serve to provide evidence-based guidance to clinicians regarding how to focus their limited time when testing is to be part of the assessment process.

Further, several respondents noted the challenge that comes from the relative lack of service availability once children enter the school years. On the one hand, identifying who is likely to need ongoing support may be less difficult than finding the resources to provide that support. On the other hand, respondents expressed discomfort with making predictions about outcomes when children who have been receiving services suddenly receive far less service in the school years. This concern aligns well with the issues that emerge from the illusory recovery hypothesis: If language learning has been accelerated via supports, but the fundamental learning challenges have not been resolved, it is reasonable to fear that gains will not be maintained once supports have been withdrawn. The overall theme that students may be falling through the cracks in the school years due to limitations in service provision is consistent with the message conveyed in Speech-Language and Audiology Canada's national campaign regarding the pressing need for school-age speech and language services, presented during Canada's most recent federal election ("Vote Communication Health Campaign;" Speech-Language & Audiology Canada, n.d.). Clinicians, researchers, and clinical associations all have a critical role to play in helping to build awareness of this need and understanding of the potential costs of these needs going unmet. Careful attention to ensuring that we are adequately identifying those children who present with lingering but potentially hidden challenges is an important piece of building that awareness.

Limitations

One goal of this paper was to stimulate reflection and discussion regarding the most appropriate course of action for children who appear to have outgrown a language disorder. While we hope that this goal was achieved, a number of limitations to the current work are important to note. With 46 respondents, the sample represents approximately 5% of those Alberta-registered S-LPs who report working primarily with children. Given this response rate, we cannot assume that the responses are broadly representative of the views and experiences of S-LPs in Alberta. We cannot confidently state why the response rate was not higher, but one contributing factor may have been the recruitment method (publishing an invitation to participate in the monthly newsletter of the provincial college). It is reasonable to think that the response rate may have been higher had invitations to participate been sent to clinicians directly. On the positive side, the respondents worked in a broad range of settings, many worked with several pediatric age groups, and the majority worked with children across the spectrum of

severity. These observations allow greater confidence in the extent to which the survey results may represent the views of S-LPs in Alberta broadly.

At the same time, for many of the questions, fewer than 46 responses were received, because not all of the questions were relevant to all of the respondents' work situations. Moreover, although we collected broad information on work context, we did not collect detailed background information such as the respondents' years of clinical experience, geographical setting (e.g., urban/rural), the proportion of children on their caseload that are multilingual, or the proportion that have additional diagnoses beyond speech and language. This information, had it been collected, would provide a better indication of the broader representativeness of the survey data.

Finally, as this survey was disseminated only within Alberta, the results reflect the views within one particular region of Canada. We suspect that the picture would not differ greatly had we surveyed within a different region of the country; however, further work is needed to confirm this view.

Conclusion

Further research on trajectories and outcomes of language disorder is important if we are to adequately advocate for and meet the needs of school-age children with DLD. Potential immediate priorities for research include documenting rates of apparent kindergarten-age resolution of DLD as well as school-age maintenance of resolved status or re-emergence of clinically significant difficulties when children are assessed using current tools and methods. A key focus of such a line of research should be to document the features of early language disorder that may predict whether normalization of language abilities and maintenance of recovered status are likely, and in particular the tools, methods or aspects of language that are likely to be sensitive to language status at kindergarten age. As Charest et al. (2019) noted, some candidate measures at kindergarten age include grammatical morphology production, sentence repetition, phonological awareness, rapid naming, narrative production/comprehension abilities, and complex syntax. Critically, such research should also begin to identify how information that goes beyond discrete test score cutoffs, such as parent and clinician perceptions, can influence our understanding of trajectories and outcomes, and how the different sources of information complement each other. Research of this scope will be best achieved through the combined, collaborative efforts of researchers and clinicians.

References

Alberta College of Speech-Language Pathologists and Audiologists. (n.d). *2014 annual report*. Retrieved from http://acslpa.ab.ca/wp-content/uploads/2015/04/2014-ACSLPA-Annual-Report1.pdf

Bishop, D. V. M., & Adams, C. (1990). A prospective study of the relationship between specific language impairment, phonological disorders and reading retardation. *Journal of Child Psychology and Psychiatry, 31,* 1027–1050. doi:10.1111/j.1469-7610.1990.tb00844

Bishop, D. V. M., & Edmundson, A. (1987). Language-impaired 4-year-olds: Distinguishing transient from persistent impairment. *Journal of Speech and Hearing Disorders, 52,* 156–173. doi:10.1044/jshd.5202.156

Bishop D. V. M., Snowling, M. J., Thompson P. A., Greenhalgh, T., & CATALISE consortium. (2016). CATALISE: A multinational and multidisciplinary delphi consensus study. Identifying language impairments in children. *PLoS ONE, 11,* 1–26. doi:10.1371/journal.pone.0158753

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3,* 77–101. doi:10.1191/1478088706qp063oa

Bzoch, K. R., League, R., & Brown, V. L. (2003). *Receptive-expressive emergent language test* (3rd ed.). Austin, TX: Pro-Ed.

Charest, M., Borger, P., Chan, C., Sanders, K., Yip, B., McFarlane, L., & Schneider, P. (2019). Assessment, diagnosis, and recovery from developmental language disorder at kindergarten age: Research review and clinical discussion. *Canadian Journal of Speech-Language Pathology and Audiology, 43,* 49–61.

Cowley, J., & Glasgow, C. (1994). *The Renfrew bus story*. Centerville, DE: The Centerville School.

Gillam, R. B., & Pearson, N. A. (2004). *The test of narrative language*. Austin, TX: Pro-Ed.

Gillam, R. B., & Pearson, N. A. (2017). *The test of narrative language* (2nd ed.). Austin, TX: Pro-Ed.

Im-Bolter, N., & Cohen, N. J. (2007). Language impairment and psychiatric comorbidities. *Pediatric Clinics of North America, 54,* 525–542. doi:10.1016/j.pcl.2007.02.008

LaParo, K. M., Justice, L., Skibbe, L. E., & Pianta, R. C. (2004). Relations among maternal, child, and demographic factors and the persistence of preschool language impairment. *American Journal of Speech-Language Pathology, 13,* 291–303. doi:10.1044/1058-0360(2004/030)

Paul, R., & Roth, F. P. (2011). Characterizing and predicting outcomes of communication delays in infants and toddlers: Implications for clinical practice. *Language, Speech, and Hearing Services in Schools, 42,* 331–340. doi:10.1044/0161-1461(2010/09-0067)

Scarborough, H. S. (2009). Connecting early language and literacy to later reading (dis)abilities: Evidence, theory, and practice. In F. Fletcher-Campbell, J. Soler, & G. Reid (Eds.), *Approaching difficulties in literacy development: Assessment, pedagogy and programmes* (pp. 23–38). Thousand Oaks, CA: Sage.

Scarborough, H. S., & Dobrich, W. (1990). Development of children with early language delay. *Journal of Speech, Language, and Hearing Research, 33,* 70–83. doi:10.1044/jshr.3301.70

Schneider, P., Dubé, R. V., & Hayward, D. (2005). *The Edmonton narrative norms instrument*. Retrieved from University of Alberta Faculty of Rehabilitation Medicine website: http://www.rehabresearch.ualberta.ca/enni

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical evaluation of language fundamentals* (4th ed.). San Antonio, TX: Pearson.

Semel, E., Wiig, E. H., & Secord, W. A. (2004). *Clinical evaluation of language fundamentals – preschool* (2nd ed.). San Antonio, TX: Pearson.

Snowling, M. J., Duff, F. J., Nash, H. M., & Hulme, C. (2016). Language profiles and literacy outcomes of children with resolving, emerging, or persisting language impairments. *Journal of Child Psychology and Psychiatry, 57,* 1360–1369. doi:10.1111/jcpp.12497

Spaulding, T. J., Plante, E., & Farinella, K. A. (2006). Eligibility criteria for language impairment: Is the low end of normal always appropriate? *Language, Speech, and Hearing Services in Schools, 37,* 61–72. doi:10.1044/0161-1461(2006/007)

Speech-Language and Audiology Canada. (n.d.). *Vote communication health.* Retrieved from http://sac-oac.ca/vote-communication-health

Stark, R. E., Bernstein, L. E., Condino, R., Bender, M., Tallal, P., & Catts, H. (1984). Four-year follow-up study of language impaired children. *Annals of Dyslexia, 34,* 49–68. doi:10.1007/BF02663613

Stothard, S. E., Snowling, M. J., Bishop, D. V. M., Chipchase, B. B., & Kaplan, C. A. (1998). Language-impaired preschoolers: A follow-up into adolescence. *Journal of Speech, Language, and Hearing Research, 41,* 407–418. doi:10.1044/jslhr.4102.407

Wiig, E. H., Semel, E., & Secord, W. A. (2013). *Clinical evaluation of language fundamentals* (5th ed.). Bloomington, MN: Pearson.

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2002). *Preschool language scales* (4th ed.). San Antonio, TX: Pearson.

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2011). *Preschool language scales* (5th ed.). San Antonio, TX: Pearson.

Authors' Note

Correspondence concerning this article should be addressed to Monique Charest, Department of Communication Sciences and Disorders, 2-70 Corbett Hall, University of Alberta, Edmonton, AB, Canada, T6G 2G4. Email: mcharest@ualberta.ca

Acknowledgments

We would like to thank Dan Moore for technical assistance with the development and management of the online version of the survey, and Julie Evans for her comments and feedback on an original draft of the survey.

Disclosures

No conflicts of interest, financial or otherwise, are declared by the authors.

Appendix

Survey Questions

1) Do you work with children?

Yes
No

Survey ends if respondent replies no

2) What age ranges do you work with? (check all that apply):

A. Toddler (1–2 years old)
B. Preschool age (3–4 years old)
C. Kindergarten age (5–6 years old)
D. Early elementary age (7–9 years old)
E. Late elementary age (9–12 years old)
F. Junior high school age and beyond (12 years+)

3) What type of setting(s) do you work in? (e.g., Community Health, Elementary School, Private Practice) ***(box for open-ended question)***

4) What populations do you work with? (check all that apply):

A. Children with mild-moderate language delays/disorders
B. Children with severe language delays/disorders
C. Children across the spectrum of severity

5) What kinds of information do you consider when making a decision about a child's diagnosis and need for intervention? (Please rank according to how heavily you typically rely on each source of information, with 1 indicating the information source that you rely on most heavily). ***(box next to each item for ranking)***

A. Standardized tests
B. Criterion referenced tests
C. Language sample analysis
D. Parent concerns
E. Teacher concerns
F. Concerns from other team members (e.g., physical or occupational therapists, nurses, physicians)
G. Clinical observations in context
H. Other (please specify)

When making observations in context, what kinds of information (such as the type of behaviours or contexts) do you find most useful? *(Open-ended question that will appear if the respondent checks off "G").*

6) If you use standardized tests, what cut-off criteria do you use (in standard deviations and/or percentile rank) for diagnosis of a language delay/disorder?

7) If you use standardized or criterion-referenced tests, please list the tests you rely on most frequently. Please list up to five in order of decreasing importance for the following age groups:

A. Preschool-aged children, up to and including pre-kindergarten (if applicable)
B. Children in kindergarten (if applicable)
C. Children in Grade 1 and beyond (if applicable)

Question 8 (a-d) asks more specifically about your experiences with children in the preschool- to school-age transition years (approximately 4–6 years of age).

8) When working with children in the 4–6 year-old age range:

8a) How confident are you in the tools that you have at your disposal to accurately identify whether or not a child has a language delay/disorder?

N/A. I do not work with children in this age range.
1 - not at all confident
2
3
4
5 - very confident

Please elaborate if desired *(space for longer answer provided):*

8b) How do you proceed if there is a discrepancy between the information provided by formal tools (standardized and/or criterion referenced tests) and your clinical judgment about diagnosis and need for intervention?

A. N/A. I do not work with children in this age range.
B. Recommend (or continue) intervention if either source of information warrants.
C. Recommend (or continue) intervention if test results warrant.
D. Recommend (or continue) intervention if clinical observations warrant.
E. Conduct further assessment.
F. Discharge client/do not recommend intervention at this time, but share any concerns with parent.
G. Discharge client/do not recommend intervention at this time, no further action.
H. Other (please specify):
Please elaborate if desired:

8c) How do you decide whether a child has recovered from an earlier diagnosis of language delay/disorder? *(box for open-ended response)*

8d) How confident are you in the tools that you have at your disposal to predict a child's risk for future communication difficulties (i.e., during the school-age years)?

N/A
1 - not at all confident
2
3
4
5 - very confident

Please elaborate if desired *(space for longer answer provided)*:

9) Are there any age ranges that you find more difficult than others to evaluate when determining whether recovery from a language delay/disorder has been achieved? If so, which ones? (Select all that apply.)

A. Toddler (1–2 years old)
B. Preschool age (3–4 years old)
C. Kindergarten age (5–6 years old)
D. Early elementary age (7–9 years old)
E. Late elementary age (9–12 years)
F. Junior high school age and beyond
G. N/A

Please elaborate if desired *(space for longer answer provided)*:

Question 10 (a, b) asks more specifically about your experience with school-age children (early elementary years and beyond).

10a) If you work with school-aged children, how often do you receive referrals for children who had previously been discharged from S-LP services?

A. Never
B. Rarely
C. Sometimes
D. Often
E. Very Often
F. I don't know/I am unsure. Please specify:
G. N/A. I do not work with children in this age range.

10b) What do you feel are the principal concern(s) that prompt referrals to S-LP services for school-age children who had previously been discharged from S-LP services? (Please rank top three, with 1 indicating the most frequent concern.) *(box next to each item for ranks)*

Concerns about:
A. Reading and written language
B. Academic achievement
C. Listening skills and/or language comprehension
D. Peer interactions
E. Speech intelligibility
F. Other (please specify)
G. N/A

In the remaining three questions, we are asking for any additional information that you would like to share about your experiences working with children as they transition from preschool-age to school-age, and making decisions about ongoing needs for support.

11) In your opinion, what resources or knowledge do we lack as a field when it comes to predicting the future risk of children with language delays/disorders diagnosed in the preschool years? *(box for open-ended response)*

12) Is there anything else that you would like to tell us about your practices or ability to predict ongoing needs through the preschool- to school-age transition (i.e., 4–6 years)? *(box for open-ended response)*

13) What other questions do you have related to the issue of diagnosis and prediction of language delays/disorders through this age range? *(box for open-ended response)*

You have reached the end of the survey. Thank you for your participation!

SAC
OAC

Speech-Language &
Audiology Canada

Orthophonie et
Audiologie Canada

Communicating care
La communication à coeur

613.567.9968
1.800.259.8519
1000-1 rue Nicholas St.
Ottawa ON K1N 7B7
www.sac-oac.ca | @SAC_OAC