

LANGUAGE DATA FROM CHILDREN WITH AUTISM SPECTRUM DISORDER:
MEASUREMENT, RELIABILITY, AND APPLICATION

By

Moon Young Savana Bak

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Special Education—Doctor of Philosophy

2019

ABSTRACT

LANGUAGE DATA FROM CHILDREN WITH AUTISM SPECTRUM DISORDER: MEASUREMENT, RELIABILITY, AND APPLICATION

By

Moon Young Savana Bak

Many interventions based on the principles of applied behavior analysis (ABA) have been regarded as evidence-based practices that are effective for children with autism spectrum disorder (ASD). Interventions based on ABA typically rely on the observation and measurement of human behavior for implementation and analysis. Thus, the strength and validity of the ABA intervention is reliant on the accurate and reliable collection of data. However, ABA researchers traditionally collect behavioral data through human observation, and this can increase the potential for error such as bias. Recent developments in technology have introduced various automated data collection apparatus that can efficiently collect reliable and more accurate data. Although other fields such as health, business, and policy have utilized automated data collection apparatus, the field of ABA has yet to fully incorporate these apparatuses to aid research. Utilizing automated data collection in ABA interventions may increase potency and inform current interventions and provide ideas for new interventions. Therefore, the current dissertation investigated the use of automated data collection in ABA research, the reliability of using an automated data collection apparatus for children with ASD, and the application of an automated data collection apparatus for language research in children with ASD in three separate but related chapters (Chapter 2, 3, and 4).

Chapter 2 investigated the current use of automated data collection apparatus in ABA research between 2010 to 2018 in a systematic literature review. A hand-search of selected ABA

journals revealed that only 149 studies out of 1466 total published studies used an automated data collection apparatus. The results provided support that ABA research may not be fully utilizing technology that can possibly increase accuracy and efficiency in measurement and data collection. Benefits and implications of using an automated data collection apparatus in the field of ABA are also discussed.

Chapter 3 provided a reliability analysis for the Language Environment Analysis (LENA[®]) system for children with ASD. The LENA system is an automated data collection apparatus developed in language studies involving typically developing children. Although many researchers use the LENA system for language research in children with ASD there has not been a stand-alone reliability analysis for this population. The primary investigator assessed the reliability of the LENA system measures by calculating the intraclass correlation coefficient (ICC) between the LENA system and the human coders. The results indicated that although the mean ICC between human coders and the LENA system was high, researchers should exercise caution when using some measures collected by the LENA system for some children with ASD.

Finally, Chapter 4 presents an exploratory analysis regarding the effects of environmental variables on the quantity of language in children with ASD. The primary investigator used the LENA system to investigate whether environmental variables such as location, instructional grouping, intervention delivery method, and learning objectives affect the quantity of child vocalization and conversational turns. Results indicated that children with ASD had a statistically significant increase in vocalizations during inclusion, in group settings, and when the intervention was delivered naturally. The implications of manipulating environmental variables to increase language teaching and learning opportunities for children with ASD are also discussed.

ACKNOWLEDGEMENTS

*The Lord has given me a well-trained tongue, that I might know how to answer the weary
a word that will waken them.*

I would like to thank my PhD advisor and dissertation committee chair, Dr. Joshua Plavnick for the encouragement, guidance, and feedback throughout my five years at Michigan State University. I would also like to thank my dissertation committee members, Dr. Emily Bouck, Dr. Matthew Brodhead, and Dr. Gary Troia. I have learned so much from you over the years. This dissertation and I would not have come this far without your support, and I am where I am because of all of you.

I would like to thank my family for their support — listening to stories that you have no context for, sending me gigantic care packages, and never asking me when I will be finished with school. Mom, dad, although you will refuse to read it, this dissertation is for you. I would also like to thank my chosen family — Yoonjung Kim and Soo Cho for giving me a place to be myself; and Youngdong Song, Yunjae Lee, and Seungmin Kim for helping me blow off the enormous amount of stress I accumulate. Thank you, professor HyungSu Kim for your sage advice.

Thank you to my doctoral comrades, Ana Dueñas (Dude, I cannot believe I am writing this section of my dissertation!), Jiyeon Park, Unhee Ju, Sarah Avendaño (It's Dexter and Willow time!), Cynde Josol (Endgame!), Emma Sipila, Courtney Maher, Hyejin Hwang, and Julie Brehmer for the feedback, laughs, and hugs only when I ask.

Finally, I would like to thank everyone who helped me conduct the three studies in this dissertation. Thank you, Gretchen Ewart, Ariel Graham, Rachel Loken, Savannah Czosky, and

Mary Garrigus for the research support. And a big thank you to everyone at the Early Learning Institute — Billy Adler, Katja Babcock, Adrea Bedard, Emily Carmody, Allison Condon, Nikki Deel, Kaylin DeHaan, Kendra Dennis, Sara Dodson, Michelle Donelson, Alexa Foote, Anjela Galimberti, Allison Germann, Zachary Huber, Darla Kril, Alison Lee, Emma Mitchell, Noel Oteto, Kathryn Reeve, Sydney Rivard, Mady Robson, Shelby Rosalik, Brittany Roulo, Matthew Rueda, Megan Rye, Greeshma Sanchula, Katelynn Sanders, Danielle Stewart, Alexandria Thomas, Erica Weber, Laura Wright, and Jiamei Zhang. This dissertation could not have happened without you all.

TABLE OF CONTENTS

LIST OF TABLES	ix
LIST OF FIGURES	x
KEY TO ABBREVIATIONS.....	xi
CHAPTER 1 Introduction.....	1
Applied Behavior Analysis Interventions	1
Automated Data Collection.....	3
Automated Data Collection in Language Research	5
CHAPTER 2 Systematic Literature Review	9
Method	13
Journal inclusion and study selection	13
Coding.....	15
Apparatus	15
Dependent variable	16
Participant and setting.....	16
Applied vs. not-applied.....	17
Inter-rater reliability.....	18
Results.....	18
Changes over time.....	19
Apparatus	20
Dependent variable	21
Settings and participants in automated data collection	22
Applied studies.....	22
Discussion	24
Limitations and future research	30
CHAPTER 3 Original Study.....	32
Language in Children with ASD.....	32
Language Assessments for Children with ASD	32
The Language Environment Analysis System	34
Reliability of the LENA System	35
The Current Study.....	37
Method	38
Participants.....	38
School-aged participants	38
Early childhood participants	39
Apparatus	41
Data collection	41
Data extraction	41

Human coding.....	42
Measures	43
Data analysis	43
Inter-rater reliability.....	43
Results.....	44
Post-hoc analysis.....	45
Discussion	47
LENA child vocalization counts	47
LENA conversational turn counts	48
Limitations and future research	50
Conclusion	50
CHAPTER 4 Original Study.....	52
Language and Children with ASD	52
Language Interventions.....	53
Measuring Environmental Effects on Language.....	55
Method	57
Participants and setting	57
Apparatus	59
Data	59
LENA data	59
Environment data	60
Procedure	61
Data collection	61
Database.....	61
Analyses.....	62
Reliability.....	63
Results.....	64
Discussion	67
Location	68
Grouping	69
Delivery method.....	70
Implications.....	70
Limitations and future research	72
Conclusion	73
CHAPTER 5 Discussion.....	75
Automated Data Collection for ABA.....	76
Reliability of Automated Data Collection Apparatuses	78
Environmental Effects on Children’s Language	81
Language Research in Children with ASD	83
APPENDICES	84
APPENDIX A Environmental Data Sheet	85
APPENDIX B Flowchart of Study Inclusion and Exclusion.....	86

REFERENCES	87
------------------	----

LIST OF TABLES

Table 1. Number of Studies that Used Automated Data Collection in Journal by Year	20
Table 2. Number of Specific Automated Apparatus Used for Data Collection by Journal....	21
Table 3. Age, MSEL Developmental Quotient, Location, and Sex of Participants	40
Table 4. Mean, range, and <i>SD</i> of ICC between human coders and the LENA system for vocalization and conversational turn counts	44
Table 5. Correlation analysis method and results between ICCs and variables.....	46
Table 6. Age Mean and Range of Adult and Child Participants	59
Table 7. The Mean, Range, and <i>SD</i> for Child Vocalization, Conversation Turn, and Adult Word Counts for Each Environmental Variable	65
Table 8. Correlational Analysis Results Between Language Measures and Environmental Variables.....	66
Table 9. Regression Coefficients from Linear Mixed Models for Child Vocalization and Adult Word Counts	67

LIST OF FIGURES

Figure 1. Time Series Graph of Percentage of Automated Data Collection Studies from 2010 to 2018.....	19
Figure 2. Graph of Number of Applied Studies and All Automated Data Studies (Except JEAB) from 2010 to 2018.....	23
Figure 3. Graph of Apparatus Used in Applied Studies and All Automated Data Studies (Except JEAB) from 2010 to 2018.....	24
Figure 4. Detailed Bar Graph of Each Participant's ICC for Child Vocalizations.....	45
Figure 5. Detailed Bar Graph of Each Participant's ICC for Conversational Turns.....	46
Figure 6. Conversational Turn Count Parameters of the LENA System	49

KEY TO ABBREVIATIONS

ABA: Applied Behavior Analysis

ASD: Autism Spectrum Disorder

BCBA: Board Certified Behavior Analyst

BI: Behavioral Interventions

BM: Behavior Modification

BT: Behavior Technician

CI: Confidence Interval

DLP: Digital Language Processor

DQ: Developmental Quotient

EIBI: Early Intensive Behavioral Intervention

IACC: Interagency Autism Coordinating Committee

ICC: Intraclass Correlation Coefficient

IRR: Inter-rater Reliability

JABA: Journal of Applied Behavior Analysis

JBE: Journal of Behavioral Education

JEAB: Journal of the Experimental Analysis of Behavior

JOBM: Journal of Organizational Behavior Management

LEAP: Learning Experiences and Alternative Program for Preschoolers and their Parents

LENA: Language Environment Analysis

LMM: Linear Mixed Model

MSEL: Mullen Scales of Early Learning

SD: Standard Deviation

SE: Standard Error

CHAPTER 1

Introduction

Autism spectrum disorder (ASD) is a developmental disorder that affects approximately one in 40 children in the United States (Kogen et al., 2018). This is a vast increase from the early 1980s when the prevalence rate was one in 5000 children (Barbaresi, Colligan, Weaver, & Katusic, 2010). Individuals with ASD generally have difficulties in social communication and functional skills, show repetitive behavior, and have restricted interests (American Psychiatric Association, 2013). Without intervention, many of these difficulties continue throughout life, negatively affecting the possibility of a meaningful independent life (Roux, Shattuck, Rast, Rava, & Anderson, 2015). However, intervention can improve developmental outcomes for children with ASD (National Research Council, 2001).

Applied Behavior Analysis Interventions

Comprehensive and focused interventions based on the principles of applied behavior analysis (ABA) are most commonly used by practitioners and parents for children with ASD (Green et al., 2006; Reichow & Wolery, 2009; Stahmer, Collings, and Palinkas, 2005). These interventions involve systematic procedures to increase or decrease targeted behavior or skill through the observation, measurement, and analysis of human behavior (Baer, Wolf, & Risley, 1968). Comprehensive interventions are a combined set of interventions that are delivered in intensive doses (e.g., 25 hours a week) over a long period of time (e.g., a year) to target overall developmental progress in individuals with ASD (Odom, Boyd, Hall, & Hume, 2010). Pivotal response training (see Koegel, Koegel, Harrower, & Carter, 1999); the Early Start Denver Model (see Estes et al., 2015); and the Young Autism Project Model (see Reichow & Wolery, 2009) are some examples of ABA-based comprehensive interventions that have produced positive

developmental outcomes in children with ASD (Dawson et al., 2010; MacDonald, Parry-Cruwys, Dupere, & Ahearn, 2014).

Unlike comprehensive interventions, focused interventions target specific behaviors or skills and are delivered for a short period of time (Odom et al., 2010). Many focused interventions based on behavior analytic principles for individuals with ASD are recognized as evidence-based practices for children with ASD (Wong et al., 2015; also see National Autism Center, 2015) and most of the comprehensive interventions are made up of several focused interventions. These focused interventions include changing the antecedent or the consequence of the target behavior for behavior change or skill acquisition. For example, antecedent procedures such as using a visual schedule can facilitate transition (Dettmer, Simpson, Myles, & Ganz, 2000) and reinforcement procedures such as the token reinforcement system can increase motivation (Matson & Boisjoli, 2009) for children with ASD in educational environments.

Despite its wide-use and effectiveness for intervention in children with ASD, there is room for improvement in the field of ABA. Applied behavior analysis involves the study of human behavior and typically relies upon human data collectors to observe and measure participant behavior (Cooper, Heron, & Heward, 2007). However, accuracy and reliability in human data collection can be threatened by observer drift (i.e., data collectors changing the predetermined definition of the target behavior over time), objectivity (e.g., data collectors may be affected by their own preconceptions of the participant), and reactivity (i.e., participant behavior affected by the presence of a human data collector; Cooper et al., 2007). Additionally, researchers may minimize the number of research participants due to the financial and temporal resources needed for data collection and analysis in using human data collection (Crowley-Koch & Van Houten, 2013).

Automated Data Collection

In recent years, many scientific fields use microprocessor-powered automated data collection apparatuses to collect and analyze data to facilitate and innovate research involving human participants (Bates, Saria, Ohno-Machado, Shah, & Escobar, 2014; Shull, Jirattigalachote, Hunt, Cutkosky, & Delp, 2014; Swan, 2013). Automated data collection can reduce potential human data collection problems in reliability, validity, and fidelity (Crowley-Koch & Van Houten, 2013). Additionally, automated data collection allows researchers to collect data continuously without human administration or oversight. The data collection apparatus will also strictly adhere to the calibrated software or algorithm without change over time. In addition, using an apparatus for automated data collection can also address fidelity issues that can occur with human data collectors who may inadvertently omit or change predetermined procedures (Gast & Ledford, 2014). Furthermore, automated data collection can allow researchers to conduct research involving a larger number of participants by reducing the resources needed for collecting and analyzing larger amounts of data (Crowley-Koch & Van Houten, 2013).

Automated data collection is used in the medical, government, and business sectors for research and implementation of various treatments and innovations regarding human behavior (Chen & Zhang, 2014; Lowe & ÓLaighin, 2014; Lyons, Lewis, Mayrsohn, & Rowland, 2014). Technological advances in accelerometers, global positioning systems, and electrocardiography allow researchers to measure human behaviors that are difficult (e.g., running distance) or impossible to see with human eyes (e.g., change in cardiovascular rhythm) using small wearable devices such as the *Fitbit* or the *Apple watch* (Lowe, & ÓLaighin, 2014). Researchers can also collect data with minimal human presence through computer-collected data from websites or pre-programmed software (Chen & Zhang, 2014).

Automated data collection has rarely been used or discussed in the field of ABA (Crowley-Koch & Van Houten, 2013; Kelly, 1977). Crowley-Koch and Van Houten (2013) published a conceptual review on automated measurement in ABA extending Kelly's (1977) literature review regarding studies published in the *Journal of Applied Behavior Analysis* that used automated data collection. In their review, Crowley-Koch and Van Houten discussed potential applications of automated data collection apparatuses that can address the limitations in human data collection. Crowley-Koch and Van Houten's review included examples such as using an eye-tracking device for an early literacy research, using computerized software to collect various data in academic content learning, using barcode scanner data from local grocery stores to study food consumption, and using a speech recognition device such as the Language Environment Analysis (LENA[®]) system in language research for children with ASD (Crowley-Koch & Van Houten, 2013).

However, even after the publication of this conceptual review by Crowley-Koch & Van Houten (2013), there has been very little empirical support for the use of automated data in ABA. Considering ABA emerged from the experimental analysis of behavior, where automated data collection was responsible for the establishment and progress of the field (see, Skinner, 1956; Springer, Brown, & Duncan, 1981), a systematic review of the current use of automated data collection may facilitate more automated data collection in ABA. Therefore, Chapter 2 of the current dissertation presents a systematic literature review that investigated the extent to which ABA research used automated data collection apparatuses.

Chapter 2 investigated whether the field of ABA utilized technological advances available to increase validity and reliability in data collection. Based on the earlier conceptual paper by Crowley-Koch & Van Houten (2013), the review evaluated behavior analytic studies

from selected ABA journals from 2010 to 2018 to identify studies that used automated data collection. Studies were coded to analyze whether automated data collection has increased since previous reviews, the automated data collection apparatus that were used, and how the apparatus were used in applied research. Specifically, the systematic literature review asked the following research questions:

1. What are the changes, if any, in the use of automated data collection over time in applied behavior analytic research?
2. What is the extent to which automated data collection is used in applied behavior analytic research?
3. What are some uses of automated data collection in applied behavior analytic research?

Automated Data Collection in Language Research

Given the evidence supporting ABA for children with ASD, utilizing automated data collection in research and interventions for children with ASD may also have added benefits, especially in areas such as language and social communication as repeated and continuous collection of data is recommended in order to assess children with ASD's language skills in the natural setting for effective intervention (Kasari, Brady, Lord, & Tager-Flusberg, 2013).

Although symptoms and severity of ASD vary greatly between individuals, difficulty in language and social communication is prevalent amongst individuals with ASD (American Psychiatric Association, 2013). Furthermore, though evidence supports both focused and comprehensive ABA-based language interventions for children with ASD (Boyd et al., 2014; Kane, Connell, & Pellecchia, 2010; Peters-Scheffer, Didden, Korzilius, & Sturmey, 2011), researchers predict at least 30% of individuals with ASD graduating from high school have minimal to little language skills (Roux et al., 2015). Language is a significant predictor for social skills development,

academic success, reducing behavioral difficulties, and quality of life in adulthood (Pickles, Anderson, & Lord, 2014; Roux et al., 2015). Consequently, the Interagency Autism Coordinating Committee (IACC) has called for increased research in language for children with ASD to decrease the number of children with ASD who have little to no language skills (IACC, 2014).

Effective and meaningful research requires valid and relevant data. Currently, many researchers use direct assessment (e.g., standardized assessment) or anecdotal reports (e.g., parent survey) for language characteristics and language intervention studies regarding children with ASD. However, direct assessments conducted during a short period of time may not be an accurate measure of the child's language skills (Luyster, Kadlec, Carter, & Tager-Flusberg, 2008; Sandbank & Yoder, 2014). In addition, anecdotal reports can be subjective depending who the reports come from (Luyster et al., 2008). Hence, researchers recommended collecting repeated (e.g., more than twice) or prolonged (e.g., longer than 30 min) natural language samples to supplement direct assessments and anecdotal reports in language studies regarding children with ASD (Bak, Plavnick, & Byrne, 2019; Sandbank & Yoder, 2014; Tager-Flusberg et al., 2009). But collecting repeated language samples requires increased time and human resources for recording and analyzing the additional language data.

Thus, several researchers have used the LENA system for automated data collection of language samples from children with ASD (e.g., Dykstra et al., 2012; Sanders et al., 2016; Warren et al., 2010). The LENA system is comprised of a software that aggregates then analyzes frequency and duration of language data, and a small (8.13 cm x 5.08 cm x 1.27 cm) digital recording device (Ford, Baer, Xu, Yapanel, & Gray, 2008). Some benefits of using the LENA system in language studies are automatic count of vocalization data for 16 hr in 5-min intervals and the opportunity to collect repeated language samples with minimal interference to

participants, especially children with ASD. Though the LENA system facilitates automated data collection of language in children with ASD, its original intention for use was for typically developing children between 12 to 48 months of age. As a result, some researchers question the reliability of using the LENA system for research in children with ASD with severe language delays or children with ASD over the age of four (see Bak et al., 2019; also, Rankine et al., 2017). Therefore Chapter 3 of the current dissertation investigated the reliability of the LENA system to support its use in language research involving children with ASD.

Chapter 3 compared data collected and disaggregated by the LENA system with data collected by human coders using natural language samples collected from children with ASD of various ages and ASD symptom severity. A preliminary study demonstrated that the reliability calculated with the intraclass correlation coefficient (ICC) between the LENA system and human coders was excellent (Cicchetti, 1994) at .87 using language samples from elementary school children with ASD (Bak et al., 2019). However, this preliminary study included a small sample (i.e., 18 hr from nine participants) from children with ASD between the ages of 6 to 10. Therefore, current study analyzed the reliability of the LENA system by calculating the ICC between the LENA system and human coders on the frequency of child vocalizations and conversation turns between the child and an adult from natural language samples randomly selected from 40 participants between the ages of 3 to 9 years.

Finally, Chapter 4 presents an exploratory study that investigated possible environmental effects on expressive language in young children with ASD using the LENA system. Previous research suggested that manipulating environmental variables may that may positively affect ongoing interventions (Boyd et al., 2014; Kane et al., 2010). Therefore, Chapter 4 investigated whether certain environmental variables could positively affect the quantity of child vocalization

and conversational turns in 21 young children with ASD in three EIBI centers. The LENA system collected and aggregated the child vocalization and conversational turn counts. These language measures were analyzed with time-matched environmental data that was collected by 32 behavior technicians in the EIBI center. The environmental variables that were of interest pertained to the location, the grouping, the delivery method, and the objective of instruction. The primary investigator conducted correlational analyses and linear mixed modeling (Raudenbush & Bryk, 2002) to investigate differences in child vocalization and conversation turn counts in children with ASD under different environmental variables within the EIBI centers. Specifically, Chapter 4 asked the following research questions:

1. Is there a difference in the quantity of selected LENA system language measures under different environmental variables?
2. Is there a correlation between selected LENA system language measures and different environmental variables?
3. And if correlations are identified, are there statistically significant differences in the quantity of selected LENA system language measures under different environmental variables?

CHAPTER 2

Systematic Literature Review

Recent advances in technology have introduced big data, various automated data collection methods, and data analytics to many research disciplines and human service industries (Bates, Saria, Ohno-Machado, Shah, & Escobar, 2014; Provost & Fawcett, 2013). In health care and health care research, for example, data are collected and stored during direct interactions between patient and doctor through electronic health records, but could also be automatically collected by cell phones, wearables (e.g., Fitbits), and social media (Bates et al., 2014). These streams of data can be useful at both the population and individual levels, and often, can include direct measures of human behavior (e.g., Shull, Jirattigalachote, Hunt, Cutkosky, & Delp, 2014; Swan, 2013). Such an approach should be of great interest to researchers and practitioners in applied behavior analysis (ABA) as it may allow for targeted interventions to improve human behavior with effective and efficient data collection (Crowley-Koch & Van Houten, 2013).

Precise measurement of observable behavior is a critical dimension of applied behavior analytic research and practice (Baer, Wolf, & Risley, 1968). Although most ABA research involves small sample sizes, behavioral research textbooks are clear that the unit of analysis, or case, can involve a large-N, possibly as large as a state if one were interested in and could reliably measure behavior at such a level (Gast & Ledford, 2014; Kazdin, 2011). It is also likely that ABA can contribute to the uses of big data and data analytics as they involve predicting and improving behavior yet are not equipped with the same behavior change tactics used in ABA interventions. A first step toward utilizing technological development regarding data collection in applied behavior analysis is to identify apparatus that can automatically collect some or all of the data necessary to improve socially relevant behaviors.

Automated measurement in ABA is not a novel concept as the experimental branch of behavior analysis was founded with the use of automated data collection (see Skinner, 1956). When automated data collection systems are used, the participant's behavior triggers the device to record the instance of the behavior, without the need for a human observer to record an instance of the event (Crowley-Koch & Van Houten, 2013; Repp & Felce, 1990). The cumulative recorder may be the most well-known behavior analytic example of automated data collection; experimental researchers tracked the lever pressing of rats or key pecks of pigeons around the clock using a device connected to an operant chamber that was engineered to mark each event on a scroll of paper each time a lever press or key peck occurred (see Skinner, 1956 for a description). A similar process can be observed in more recent experimental research whereby participants complete tasks on computers, with computer software tracking participant responses under varied presentation of stimuli (e.g., Finn, Barnes-Holmes, Hussey, & Graddy, 2016). The goal of these devices is to remove as much experimenter bias from the measurement process as possible and to obtain highly precise recordings of human behavior under specific environmental conditions (Crowley-Koch & Van Houten, 2013).

The importance of measurement in ABA is very similar to the experimental arm of behavior analysis (Springer, Brown, & Duncan, 1981). However, experimental studies are often interested in easily assessed and quantified behavior, such that measurement can be conducted with automated devices. The applied branch of behavior analysis focuses on behavior change that “enhance and improve people’s lives” and chooses behaviors that are “socially significant for participants” (Cooper, Heron, & Heward, 2007, p. 16). Consequently, the complexity of human behavior as a dependent variable in applied behavior analysis has historically required human observation and has limited the extent to which variables could be automatically recorded

(Kelly, 1977). An advantage of automated data collection, such as an operant chamber, is the precision of measurement across smaller temporal units and for longer periods of time relative to the capacity of human observers, thereby providing precision and detail that might not be possible or financially feasible with human observation (Crowley-Koch & Van Houten, 2013). For example, covert behaviors such as anxiety, fear, arousal have been traditionally measured through human observation sometimes supplemented by participant surveys (Kazdin, 1979; Romanczyk, Kent, Diament, & O’Leary, 1973). Advances in technology bring new potential to automatic data collection in ABA and might offer opportunities to quantify variables that are important to further the applied science.

Recent advances in microprocessor technology offer many opportunities for scientists to study a wide range of human behavior in applied settings (Lowe, & ÓLaighin, 2014; Bonato, 2005; Lyons, Lewis, Mayrsohn, & Rowland, 2014). These advances present an opportunity for applied behavior analytic researchers to collect more precise and comprehensive data on dependent measures of interest (Crowley-Koch & Van Houten, 2013), which may expand the reach of ABA within other areas of human behavioral science (e.g., education, health policy).

Kelly (1977) systematically evaluated the quality of data collection in applied behavior analytic research by documenting the extent to which studies published during the first eight years of the *Journal of Applied Behavior Analysis* (1968-1975) ensured reliability and validity of measurement of the dependent variable. Among other variables, Kelly (1977) assessed the number of applied studies in the journal that incorporated automated data collection within the procedures. Kelly reported that 16% of studies published in the review timeframe used automated data collection. At that time, the complexity of the dependent variable in ABA was believed to necessitate human observation and limited the extent to which variables could be

automatically recorded. In addition, applied environments were not readily turned into operant chambers where subjects' behavior could be automatically recorded by devices within that environment.

Although modern technology has introduced numerous digital devices that expand the behaviors and subjects to which automated data systems can be used, applied behavioral researchers may not be taking advantage of the technology and the benefits of automated data collection (Crowley-Koch & Van Houten, 2013). In the time since Kelly's (1977) review, we know of no systematic reviews of the extant literature evaluating the use of automated data collections in applied behavior analytic research.

Crowley-Koch and Van Houten (2013) suggest the lack of emphasis on automated data is a barrier to the widespread adoption of ABA. They provide an overview of potential devices that automatically collect and extract data for a number of dependent variables that may be of interest to behavior analysts (e.g., language, location, movement). Although their commentary provides several potential systems for automatically collecting data in applied behavioral research, Crowley-Koch and Van Houten (2013) only estimate the extent to which such applications are currently in use.

An updated systematic review has potential to advance current data collection practices in applied behavior analytic research by identifying devices that have been successfully used to automatically collect data, as well as those that have potential for broader adoption. Therefore, the current review evaluated the extent to which influential ABA journals published studies that used automated data and the type of automated data collection apparatus these studies used for the past nine years. This review aims to extend measurement practices in behavior analysis by systematically evaluating the type and frequency of automated data collection and offers

empirical support to the recommendations of Crowley-Koch and Van Houten (2013). As a comparison, the current review also investigated the use of automated data collection during one publication year in the *Journal of the Experimental Analysis of Behavior* (JEAB). This allowed for a comparison of automated data collection in ABA research to current uses in the experimental arm of behavior analysis. Specifically, the current review asks the following research questions:

1. What are the changes, if any, in the use of automated data collection over time in applied behavior analytic research?
2. What is the extent to which automated data collection is used in applied behavior analytic research?
3. What are some uses of automated data collection in applied behavior analytic research?

Method

Journal inclusion and study selection. Journals were included in the present review if the journal (a) was published in English, (b) was included in the *Social Sciences Citation Index Master Journal List* from *Clarivate Analytics* (see <http://mjl.clarivate.com/cgi-bin/jrnlst/jloptions.cgi?PC=SS>) and stated in their publication statement or overview that the journal publishes studies based on applied behavior analysis or behavioral sciences at the time the review was conducted, and (c) regularly published at least 50% of applied research examining human behavior among all experimental articles during the most recent previous year (i.e., 2017). The journal list from *Clarivate Analytics* was used as the database provided the largest list of peer-reviewed journals compiled by a third-party ([https:// clarivate.com](https://clarivate.com)). Journals selected for this review according to these criteria were: *Behavioral Interventions* (BI), *Behavior Modification* (BM), *Journal of Applied Behavior Analysis* (JABA), *Child and Family Behavior*

Therapy, Journal of Behavioral Education (JBE), Journal of Organizational Behavior Management (JOBM), and Journal of Positive Behavior Interventions.

From the included journals, research studies were selected using a two-step process. First, the authors collected all articles from each of the journals listed above that included at least one data-based study published between January 2010 and December 2018. This time period was selected because it was the most current decade and because the proliferation of consumer-products that measure biometric information happened during this decade. Data-based studies were defined as research articles or brief reports that involved collection and analysis of behavioral data. Literature reviews, conceptual articles, tributes, and book reviews were excluded. In this initial collection, the authors also included all articles with at least one data-based study from JEAB in 2015 to analyze possible differences and or similarities in the use of automated-data collection between applied and experimental fields of behavior analysis.

The researchers then reviewed all articles to determine whether the initial articles met the following criteria for final inclusion in the review: (a) the dependent variable was an output collected and displayed by an automated data-collecting apparatus, (b) this apparatus operated on a pre-calibrated or pre-programmed algorithm specific to the dependent variable, and (c) the data displayed by the apparatus was in the same form as the data the researchers reported in the manuscript. For example, office referrals collected and entered into a school wide behavior intervention software did not meet criteria for automated data collection because the software required participants or investigators to observe, assess, and code behavioral events and directly enter data into the software (i.e., the apparatus in this case simply was used for organization and management and not for automated data collection). Similarly, studies that used electronically collected surveys were not included in the review because although the survey aggregated the

results for the researchers, the participants had to manually input the answers into the survey. However, if an apparatus was used to determine and numerically display the insulin levels of participants, the study was coded as using automated data, as this apparatus operated on an algorithm specific to collect insulin levels and investigators only had to retrieve the data from the apparatus. If an article included multiple research studies, each study was assessed individually for inclusion in the review. The review process produced 17 studies from BI, seven studies from BM, 96 studies from JABA, four studies from JBE, and 26 studies from JOBM. Both *Child and Family Behavior Therapy* and *Journal of Positive Behavior Interventions* did not have any studies that used automated data collection within the timeframe (i.e., 2010 to 2018) for this review. Additionally, the authors identified 54 studies from the 2015 issue of JEAB. A total of 204 studies were included for coding and analysis. A flowchart depicting the process of inclusion and exclusion of studies is presented in Appendix A.

Coding. Studies that were included in the review were coded by the authors for the following descriptive components of each study: apparatus, dependent variable, participant, and setting. All coders were trained to 90% inter-rater reliability (IRR) with the second author on articles from the 2009 volume of JABA prior to coding studies included in this review. Coders received a training study, independently coded the study for each dependent variable, and compared results with the second author. Disagreements were discussed, and coders completed another training-study until reaching 90% IRR on a single study. The IRR percentage was calculated by dividing by the sum of agreed variables by the total number of variables and multiplying it by 100 to yield a percentage (Gast & Ledford, 2014).

Apparatus. The authors classified the apparatus used in each study within one of six categories (described below). The type of apparatus included computer, wearable, sensor,

counter, laser, custom-built, or other. Studies that used a computer running a specific software or program to measure and collect the data were coded as, computer. However, a study that used a computer only to control an externally connected apparatus that directly measured and collected the dependent variable was coded according to the connected apparatus. For example, a barcode reader that needed to be connected to a computer for calibration and data input/output was coded as, laser (e.g., Sigurdsson, Larsen, & Gunnarsson, 2014). A wearable apparatus was defined as any digital apparatus worn by a participant that recorded and displayed physical movement, such as a *Fitbit*. Sensors included any apparatus that detected and measured an organic chemical reaction brought on by physiological changes in participants such as breathalyzers that detect alcohol or oxygen levels. Counters included any apparatus that tabulated the frequency of an event or a behavior such as a pedometer. Laser was defined as an apparatus that incorporated lasers to detect and measure the dependent variable such as barcode readers and speedometers. Apparatuses that included any custom-built apparatus such as an operant chamber that could track the response to a behavior of interest were coded as, custom-built (see Skinner, 1956). Studies that used automated data collection with devices that did not fit into any of the above categories were coded as other.

Dependent variable. Authors coded two dimensions of the dependent variable. First, authors coded how the dependent variable was collected in each study. A study was coded as *apparatus only* if the only measure of behavior in the study was collected solely by an apparatus. If there were additional dependent variables collected for the study, and these were collected in-part by humans, the study was coded as *apparatus and human*.

Participant and setting. The type of participant in each study was coded into one of the following four categories: animal, university students, adult community members, and minor

community members. Studies that used animals instead of human participants were coded as animals. University students were defined as undergraduate or graduate students enrolled in a higher education institution who volunteered or were recruited specifically for the study. A likely representation of this group would be graduate or undergraduate students attending the university where the study was conducted. Adult community members were defined as individuals above 18 years of age who participated in the study because they belonged to a specific community of interest (e.g., medical diagnosis, occupation) to the study. Minor community members included participants who were under 18 years old or were between grades pre-k to 12 to include individuals with disabilities still receiving K-12 education even after reaching the age of 18.

For the setting variable, the authors referenced the method section of each study and recorded the location where each study was conducted (i.e., the location where the apparatus for data collection was used).

Applied vs. not-applied. Authors evaluated each study to code whether the data collection apparatus was used for applied or experimental research. The rationale for coding the articles on whether they were applied research, was to answer the research questions on the extent to which automated data collection is used in applied research and to provide suggestions for the use of automated data collection apparatus in ABA research. The authors referenced each study's research question and the coded results for *participant and setting* (explained above) to code for this variable. In the scope of this review, for a study to be considered as applied research, the research questions had to be socially valid (Cooper et al., 2007; Wolf, 1978) for the participants of the study (e.g., gamblers in a gambling addiction study but not undergraduate volunteers with no gambling history in a gambling addiction study) and the participants had to have a "close relationship" with the setting (e.g., individuals with gambling addiction in a casino but not

undergraduate students with no gambling addiction history in a casino; Baer et al., 1968, p. 92). If studies involving animals were collected, they were coded as either applied or not-applied research with the same process. For example, a study that evaluated different training procedures for police dogs to detect narcotics would be coded as applied, but a study that used rats to test potential punishment strategies were coded as not-applied.

Inter-rater reliability. Independent reviewers assessed IRR on two levels. First, the first author selected a year between 2010 and 2018 randomly for each journal using Google's random-number generator (i.e., BI = 2015, BM = 2017, JABA = 2011, JBE = 2011, JOBM = 2016). The first author and an independent reviewer selected studies that met the inclusion criteria for automated-data use from each journal volume. The planned IRR for article inclusion was to be calculated using the intraclass correlation coefficient. However, the first author and independent reviewer had identical selections for each journal volume.

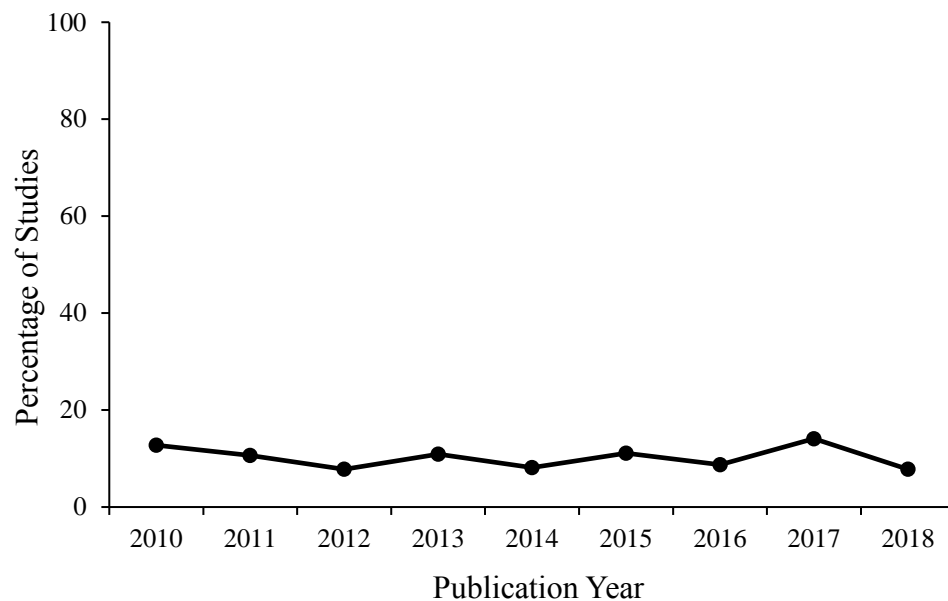
Inter-rater reliability was also calculated for each of the dependent variables among the included studies. IRR between the first author and four independent reviewers was assessed for this part of the coding process. After meeting training criteria (i.e., 90% agreement), the independent reviewers coded 33% of all included studies randomly selected across the different journals. The studies included in the IRR process contained six articles from BI, three articles from BM, 31 articles from JABA, one article from JBE, nine articles from JOBM, and 18 articles from JEAB. The mean IRR for coding was calculated by dividing the number of agreements with the sum of agreements and disagreements and by multiplying the result with 100 to derive a percentage. The IRR was 99.08% for all included articles with a range of 75 to 100.

Results

Of the 1466 data-based articles published in BI, BM, JABA, JBE, and JOBM between

2010 to 2018, 10.16% (n = 149) used an automated data collection apparatus. The *Journal of Organizational Behavior Management* had the highest (i.e., 22.81%) and JBE had the lowest (i.e., 2.4%) percentage of studies that used automated data collection during the nine years included in this review. Conversely, 71.43% (n = 35) of data-based articles published in JEAB during 2015 used an automated apparatus for data collection.

Figure 1. Time Series Graph of Percentage of Automated Data Collection Studies from 2010 to 2018



Line graph depicting the percentage of studies that used an automated data collection device during each year from all journals included in this review except JEAB.

Changes over time. Figure 1 represents a time series graph of the total percentage of studies that used automated data collection in each year from 2010 to 2018. Upon visual analysis, there were no notable increases or decreases in the percentage of published studies that used automated data collection between 2010 to 2018. By year, the highest percentage of studies that used automated data collection was in 2017 with 14.1% (n = 22 out of 156 studies) and the lowest was in 2012 and 2018, both at 7.78% (n = 13 out of 167 studies). Furthermore, there was no notable increase or decrease within each journal. Although BI published six studies that used

automated data collection in 2018 (a three-fold increase from 2017), the journal had a decline in the number of published studies that used automated data collection since 2010 ($n = 3$) with no studies using automated data collection during both 2015 and 2016. Other journals showed a similar pattern, where the numbers of published studies that used automated data collection fluctuated throughout the years included in this review. Table 1 shows the number of studies that used automated data collection in each journal by year.

Table 1. Number of Studies that Used Automated Data Collection in Journal by Year

	BI	BM	JABA	JBE	JOBM	Total
2010	3	0	14	0	3	20
2011	3	1	12	1	2	19
2012	1	0	8	0	4	13
2013	1	0	10	1	6	18
2014	1	1	10	0	2	14
2015	0	0	12	0	4	16
2016	0	1	11	1	1	14
2017	2	4	14	1	1	22
2018	6	0	4	0	3	13
Total	17	7	95	4	26	149

Note. BI = Behavioral Interventions, BM = Behavior Modification, JABA = Journal of Applied Behavior Analysis, JBE = Journal of Behavioral Education, JOBM = Journal of Organizational Behavior Management

Apparatus. Similar to the number of published studies that used automated data collection from 2010 to 2018, there was no noticeable increase or decrease in the number of apparatus used between the years. Among the different types of apparatus coded for this review, researchers used the computer most often at 59.38% ($n = 95$). Studies that used computers utilized programmable software such as *Microsoft's Visual Basic* (e.g., Fineup, Covey, & Critchfield, 2010; Fahmie, Macaskill, Kazemi, & Elmer, 2018), pre-programmed computer

software to collect the dependent variable automatically (e.g., Grindle, Hughes, Saville, Huxley, & Hastings, 2013; Schnell, Sidener, Debar, Vladescu, & Kang, 2018), or used computers as a means to control microswitches or external mechanisms (e.g., Kelley, Liddon, Ribeiro, Grief, & Podlesnik, 2015; Lancioni et al., 2017). The second most used apparatus was wearable biometric devices such as Fitbits at 10.63% (n = 17). The least used apparatus for automated data collection was the laser (e.g., speed gun; Vanwagner, Van Houten, & Betts, 2011) and apparatus in the “other” category, both at 3.13% (n = 5). All five studies that used apparatuses coded as “other” used a scale to measure weight (e.g., Darling, Fahrenkamp, Wilson, Karazsia, & Sato, 2017; Hankla, Kohn, & Normand, 2018; Napolitano, Lloyd-Richardson, Fava, & Marcus, 2011; Penrod, Wallace, Reagon, Betz, & Higbee, 2010; Peterson, Piazza, & Volkert, 2016).

Table 2. Number of Specific Automated Apparatus Used for Data Collection by Journal

	Computer	Wearable	Sensor	Counter	Laser	Custom	Other
BI	7	3	0	5	1	1	2
BM	3	2	3	1	0	1	2
JABA	63	11	8	9	2	5	1
JBE	3	1	0	0	0	0	0
JOBM	19	0	5	0	2	0	0
Total	95	17	16	15	5	7	5
JEAB	25	0	2	0	0	27	0

Note. BI = Behavioral Interventions, BM = Behavior Modification, JABA = Journal of Applied Behavior Analysis, JBE = Journal of Behavioral Education, JOBM = Journal of Organizational Behavior Management; Total is the number of studies from BI, BM, JABA, JBE, and JOBM

Dependent variable. Of the 149 studies collected in this review, 81.21% (n = 121) collected their dependent variables solely with an automated data collection apparatus. The remaining 18.8% (n = 28) used human data collectors to collect additional dependent variables in

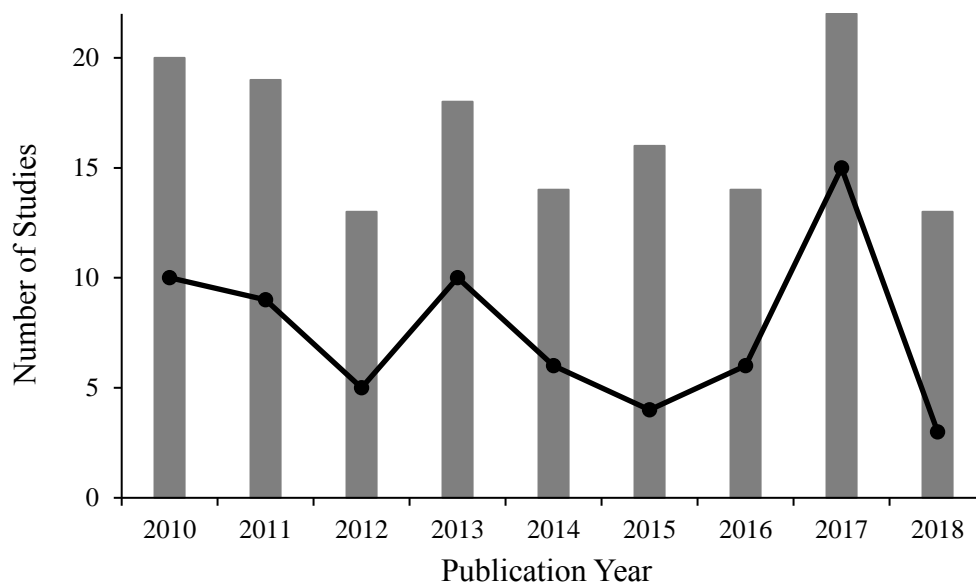
addition to the dependent variable collected by an automated apparatus. These included standardized assessment results that would accompany the dependent variable collected through an apparatus (e.g., Storey, McDowell, & Leslie, 2017); additional surveys and self-reports collected from the participants (e.g., McLeish, Luberto, & O' Bryan, 2016); and data from certain participants that could not be collected solely from an automated apparatus (e.g., Saini, Fisher, & Pisman, 2017). Similarly, 14.89% ($n = 7$) of articles coded from the 2015 issue of JEAB used human data collectors to collect additional dependent variables.

Settings and participants in automated data collection. Researchers conducted 55.03% ($n = 82$) of the studies in a not-applied setting such as a university laboratory and 44.97% ($n = 67$) in an applied setting such as public schools (e.g., Plavnick, Thompson, Englert, Mariage, & Johnson, 2016) and playgrounds (e.g., Galbraith & Normand, 2017); community establishments such as bars (e.g., Kazbour & Bailey, 2010) and local grocery stores (e.g., Sigurdsson et al., 2014); and public areas such as roads (e.g., Dixon et al., 2014). One study did not provide specific details to the setting but was coded as conducted in a not-applied setting based on the methods presented in the article (see, Mahon, Lyddy, & Barnes-Holmes, 2010). The majority of the studies (i.e., 62.42%; $n = 93$) recruited participants that were directly related to the research question such as children with cochlear implants (e.g., Golfeto & de Souza, 2015), adults with alcohol dependence (e.g., McDonell et al., 2012), and adults with Alzheimer's (e.g., Steingrimsdottir & Arntzen, 2011). A few studies recruited large scale participants such as shoppers at a local grocery store (e.g., Sigurdsson et al., 2014) or all households in a designated area (e.g., Oliveira-Castro, Foxall, & Wells, 2010).

Applied studies. Based on the coding results from the setting and the participants presented above, 30.77% ($n = 68$) studies were coded as being applied studies. Figure 2 depicts a

line graph that shows the number of applied studies that used automated data collection over bars that represent the number of all studies that used automated data collection that were published in BI, BM, JABA, JBE, and JOBM. Similar to the change in the overall number of published studies that used automated data collection, the numbers fluctuate between the years but there is no noticeable increase or decrease throughout the nine years in the number of applied studies. Per journal, 76.47% (n = 13) of studies that used automated data collection from BI were coded as applied, followed by 75% (n = 3) from JBE, 57.14 % (n = 4) from BM. There were less applied studies amongst those that used automated data collection in JABA at 42.1% (n = 40) and JOBM at 30.77% (n = 8). Only one study (see Ribeiro, Miguel, & Goyos, 2015) from JEAB was coded as an applied study of the 54 studies that used automated data collection in 2015.

Figure 2. Graph of Number of Applied Studies and All Automated Data Studies (Except JEAB) from 2010 to 2018

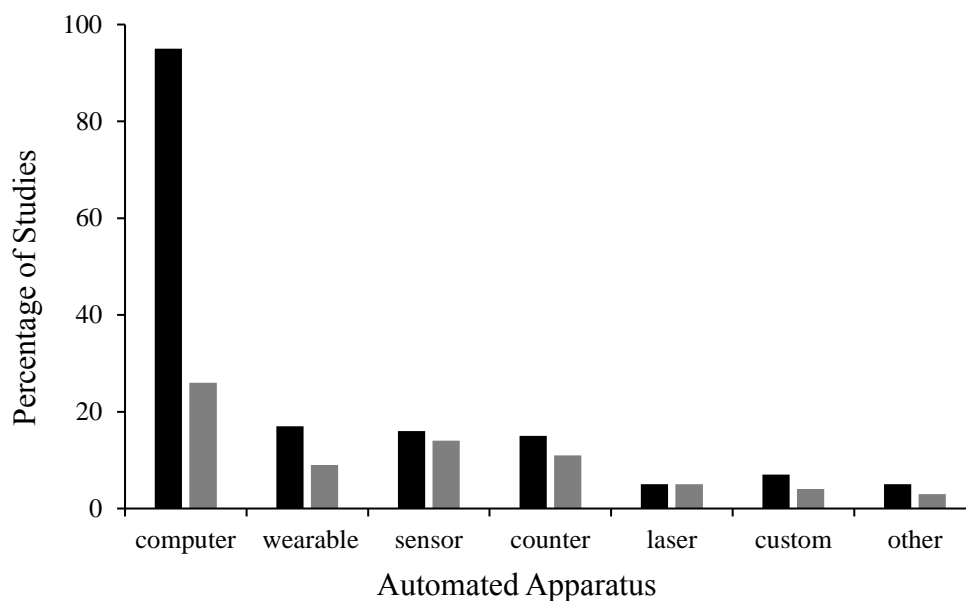


Bar graph represents the number of studies per year from all collected studies except JEAB that published research that used an automated data collection device and the line graph represents the number of applied studies that used automated data collection published each year from all collected studies except JEAB.

Although computers were still the most used automated data collection apparatus amongst applied studies, only 36.11% (n = 26) of applied studies used computers compared to

59.38% for all studies, both applied and not-applied. A larger ratio of sensors, wearables, and counters were used in applied research compared to all studies, both applied and not-applied. Sensors were used in 19.44% ($n = 14$) of applied studies compared to 10% for all studies, wearables were used for automated data collection for 12.5% ($n = 9$) of the applied studies compared to 10.63% for all studies, and counters were used in 15.28% ($n = 11$) of the applied studies compared to 9.38% for all studies, both applied and not-applied.

Figure 3. Graph of Apparatus Used in Applied Studies and All Automated Data Studies (Except JEAB) from 2010 to 2018



A bar graph of the different automated data collection apparatus used by studies included in the review. The black bars represent the percentage of each apparatus used in all automated data collection studies collected (except JEAB) and the grey bars represent the percentage of each apparatus used in applied studies only.

Discussion

The current study reviewed the extent to which automated data collection is used in studies published in behavioral research journals from 2010 to 2018. A total of 149 studies from five journals (i.e., BI, BM, JABA, JBE, and JOBM) that used automated data collection were included in this review. Of all published data-based studies across the nine years from the five journals, 10.16% collected data automatically with a range of 7.78% to 14.1% between journals.

However, there was no noticeable change over time in the percentage of studies that used automated data collection in all journals from 2010 to 2018.

Although researchers can collect data accurately, efficiently, and across a wider range of human behavior in social science, health, medical, consumer analytics, and policy through the development of technology (Bonato, 2005; Lyons et al., 2014), there was no increase in the percentage of automated data collection reported from 2010 to 2018 in ABA based on the studies collected for this review. The results of this review strongly support Crowley-Koch and Van Houten's (2013) view that the field of ABA is not thoroughly utilizing available technology that can innovate data collection methods and increase efficiency and accuracy through automated data collection. Moreover, a previous review conducted by Kelly (1977) reported 16% of studies used automated data collection from one ABA journal (i.e., JABA) between 1968-1975. The percentage of studies that used automated data collection published in JABA between 2010-2018 was 13.25%. Because the percentage of published studies that used an automated data collection apparatus in JABA from 2010-2018 is smaller than Kelly's 1977 review result, this may point to a decrease in studies that use automated data collection apparatuses since the early years of ABA. Also, the total number of published studies in JABA during the time period of the current review is much higher than the number of published studies during the time period of Kelly's (1977) review. It is also worth noting that although the field grew and more research is conducted, the ratio of automated data collection has not increased.

This lack of change is surprising considering the abundance of computers and devices that offer wireless and microprocessor technology in the current decade (Lowe, & ÓLaighin, 2014; Bonato, 2005; Lyons, et al., 2014). The foremost benefit of utilizing an automated data collection apparatus for applied behavior analytic research can be increased accuracy in

measurement (Crowley-Koch & Van Houten, 2013). Although human observation of behavior has been the traditional approach to data collection in ABA (Springer et al., 1981), human data collection can result in unintended errors (Gast & Ledford, 2014). Human observers can be affected by observer drift which require assessment in procedural fidelity and training. Human observers can also be subjected to bias, and continued observation sessions with multiple dependent variables or multiple participants can result in errors due to fatigue and stress (Gast & Ledford, 2014; Kazdin, 2011). An automated data collection apparatus can offer researchers with a data collection method of consistency and accuracy (Crowley-Koch & Van Houten, 2013).

For example, Yu, Moon, Oah, and Lee (2013) installed weight sensors in participants' chairs to measure appropriate sitting postures during sedentary office work under different feedback conditions. Had Yu and colleagues (2013) used human observers, the observers would have had to record the shift in participants' posture continuously to collect the dependent variable and report the changes immediately to allow respective feedback conditions or provide feedback accurately themselves. Yu and colleagues (2013) defined appropriate posture with definitions for five different body parts in their study. Observing participants while remembering the definitions for appropriate behavior, collecting data, and providing intervention would be a difficult task for humans that can increase errors. However, by inserting several sensors into the chairs in the participants' natural work space, Yu and colleagues (2013) were able to not only collect accurate data but also provide accurate feedback based on the parameters of the intervention (i.e., immediate vs delayed feedback).

With the advent of technology, various devices equipped with pre-programmed algorithms to measure and regulate behavior have been available for mass consumption. Researchers can use these devices as automated data collection apparatus to answer research

questions in place of human observers. Wearable devices such as the *Fitbit* can provide frequency or distance of movement for studies involving physical interventions (e.g., Washington, Banna, & Gibson, 2016) or heart rates for studies involving intensity of physical activity (e.g., Larson, Normand, & Hustyi, 2011) or anxiety levels (e.g., Chok, Demanche, Kennedy & Studer, 2010). Although some of these studies may have limitations regarding the validity of the data (i.e., using number of steps for exercise or heart rates for anxiety levels), they provide some preliminary ideas in ways an automated apparatus can increase accuracy in data collection in applied settings compared to direct human observation.

In addition to providing a method for accurate and consistent data collection, automated data collection can offer efficient and effective ways to collect behavioral data that could be considered difficult to collect by human observation (Crowley-Koch & Van Houten, 2013). For example, Reyes, Vollmer, and Hall (2011) used a penile strain gauge to discern sexual arousal among sex offender participants with developmental disabilities. The penile plethysmograph allowed Reyes and colleagues (2011) to collect data on sexual preferences through an ethical and comprehensible method for the participants (all participants were deemed “incompetent for trial” because of their disabilities) within the settings the participants were confined (i.e., a residential treatment facility). As such, the results of Reyes and colleagues (2011) study could offer solutions for intervention and treatment based on reliable, accurate data that may not have been possible without the automated data collection apparatus.

Another example is Dallery, Raiff, and Grabinski’s (2013) study where 77 adult participants recorded their own carbon-monoxide levels in their homes using a carbon-monoxide monitor and transferred the data to the researchers via the internet for a study regarding the effect of feedback methods on nicotine addiction. Dallery and colleagues’ (2013) use of an automated

data collection apparatus possibly reduced errors (e.g., counting or reporting the number of cigarettes consumed from 77 participants); logistical constraints (e.g., pre-setting a time of day for observation for the large number of participants); and potential reactivity (e.g., smoking less during the observation sessions) that could have existed with human observation (Crowley-Koch & Van Houten, 2013).

Furthermore, technological development is increasingly offering smaller microprocessors with enough processing power for continuous analysis of complex algorithms involving human movement (see Lowe & ÓLaighin, 2014). In one example, Lancioni and colleagues (2011) used camera-based microswitches to track facial movements in children with disabilities. This method can be adapted and applied to the abundance of existing studies in ABA regarding children with severe disabilities and social skill development such as eye-gaze and joint-attention. Where researchers have traditionally used human observation whether in vivo or through videotaped records to count instances of “eye contact” or facial expressions such as smiling, a programmed microswitch connected to a camera can collect minute changes in facial expression with consistent accuracy for as long as the researchers desire.

In another example, Saunders and Saunders (2012) programmed their microswitch for adults with severe disabilities so they could learn to control devices such as digital music player with their switches. The microswitch allowed the participants with severe disabilities to control preferred reinforcement devices and also allowed the researchers to automatically collect the frequency and duration of the participants’ microswitch usage (see Saunders & Saunders, 2012). The microswitch used in Saunders and Saunders (2012) can be thought of as a modern-day cumulative recorder (see Skinner, 1956) and a good example of how technology has progressed automated measurement in ABA to allow precise data collection and intervention that would

have been difficult with human observation.

Automated data collection can also allow ABA research to increase the number of participants enrolled in studies. The precision of measurement and reliability in employing a large number of human observers has sometimes limited implementation to a manageable sample size (Gast & Ledford, 2014; Kazdin, 2011). Obviously, research involving individualized instruction or those that target specific populations do not need a large sample such as interventions that target self-injurious behavior in individuals with severe disabilities (see Matson & LoVullo, 2008). However, using automated data collection apparatuses may allow ABA researchers to conduct behavior change intervention research for a larger population. Incorporating automated data collection to conduct studies with a larger sample may increase collaboration with data analytic research conducted in other fields that may be familiar with analyzing big data results (see Bates et al., 2014; also, Provost & Fawcett, 2013) but not equipped with the effective behavior change interventions that ABA can offer (Crowley-Koch & Van Houten, 2013).

Typically, behavioral studies relied on computers running programmed interfaces to simulate applied settings or offer efficient data collection for a large number of participants (e.g., Fineup et al., 2010; Hirst, DiGennaro Reed, & Reed, 2013; Roose & Williams, 2018; Tanji & Noro, 2011). As such, computers were the most used automated data collection apparatus in the current review. Computers running program language such as Visual Basic (e.g., Roose & Williams, 2018; Tanji & Noro, 2011), interactive software such as Adobe Captivate (e.g., Schnell et al., 2018), or software developed for e-learning (e.g., Critchfield, 2014; Jamison et al., 2014) can offer an advantage to researchers because it allows them to simulate applied settings or offer efficient data collection from a large number of participants.

However, many of the studies that used a computer for automated data collection were not considered as applied studies in the scope of the current review as they simulated environments or used volunteers as participants (e.g., university students). One idea for increasing the number of participants using automated data collection in applied settings is to use apparatus that were already implemented by existing infrastructure to conduct studies intended for a larger sample. Bekker and colleagues (2010) and Schultz, Kohn, and Musto (2017) used electricity meters installed in university dorms to collect dependent variables for interventions on energy conservation. In these studies, the researchers used data collected automatically by existing apparatuses (i.e., electricity meters installed by power companies) instead of observing changes in behavior such as turning off un-used devices or reducing the duration of a shower by visiting participants' residences for observation sessions (Bekker et al., 2010; Schultz et al., 2017). Using this existing automated data collection apparatus, Bekker and colleagues (2010) implemented their interventions on 326 participants and Schultz and colleagues (2017) implemented their intervention on 99 participants. Similarly, Sigurdsson and colleagues (2014) studied the effects of advertisement and in-store placement on changed purchasing habits by using the data already collected from the stores through the installed barcode scanners for checkout. With the existing apparatus, Sigurdsson and colleagues (2014) were able to enroll 100,000 participants in their alternating treatments design study.

Limitations and future research. Although this review presents quantified and systematic support to Crowley-Koch and Van Houten's (2013) review and extends Kelly's (1977) review, there are some limitations. First, we only used one journal index (i.e., Clarivate Analytics) to provide parameters for the hand search. Although Clarivate Analytics is the largest journal citation index (see <https://clarivate.com/>), we may have missed other behavioral journals

that were not included in this index. Future studies should combine a hand search of relevant journals with a database search with search terms related to ABA in order to expand the review.

Second, the current review was limited to studies published between 2010 to 2018. The apparatuses collected in the current review are heavily concentrated on personal computers and consumer biometric devices that became available for mass consumption during the current decade. Including studies that were published in the late 1980s or the early 1990s may provide innovative ideas of automated data collection that do not rely on high-technology consumer products but rather required researchers to create custom-made devices that automatically collected data – perhaps a device that bridges the cumulative recorder (see Skinner, 1956) and the microprocessor (see Saunders & Saunders, 2012).

Finally, because we only collected articles from one year (i.e., 2015) of JEAB, the current review was not able to provide a more thorough analysis of the difference in automated data collection apparatus usage between the applied and the experimental arm of behavior analysis. Future studies should include studies published throughout the years in JEAB and other experimental behavior analysis journals. This would allow for a more comprehensive overview on the automated data collection apparatuses used in not-applied studies that could lead to automated data collection apparatuses and methods that can be adapted for applied research not discussed in the current review.

CHAPTER 3

Original Study

Language in Children with ASD

Atypical language development and difficulty in social communication are common characteristics among individuals with autism spectrum disorder (American Psychiatric Association, 2013). Language is an important skill that can help ameliorate problematic behaviors (e.g., self-injurious behavior), increase social skills, and facilitate academic success (Pickles, Anderson, & Lord, 2014). However, many children with autism spectrum disorder (ASD) who do not acquire functional communication by age 5 may remain nonverbal throughout their lifetime (Tager-Flusberg & Kasari, 2013). Thus, the characteristics of language development and language skills among children with ASD are frequently examined in research as they can inform more effective interventions that promote better developmental outcomes (Sandbank et al., 2017). Despite the importance of precise measurement of language, there are many barriers to obtaining reliable language measures of children with ASD.

Language Assessments for Children with ASD

Language assessments for children with ASD in research often involve direct standardized assessments, parent report, or analysis of language samples (Luyster, Kadlec, Carter, & Tager-Flusberg, 2008; Sandbank & Yoder, 2014). Direct standardized assessment typically involves a trained assessor using manualized testing procedures to assess expressive and receptive language. The Mullen Scales of Early Learning (MSEL; Mullen, 1995) and the Preschool Language Scale (Zimmerman, Steiner, & Pond, 2002) are common direct standardized assessments utilized by researchers to evaluate language skills in children with ASD (e.g., Baril & Humphreys, 2017; Boyd et al., 2014; Weismer, Lord, & Esler, 2010). However, because

standardized assessments are usually conducted on a single day, the resulting score may not provide an accurate representation of a child with ASD's language skills which can be affected by the child's physical condition at the time of assessment, and the familiarity of the assessor and/or the assessment setting to the child (Sandbank & Yoder, 2014). In addition, for some children with ASD, standardized assessments may not produce scores that can provide sensitive results when the majority of the questions are too difficult for the child (Rankine et al., 2017).

Parent-reported language assessments are typically conducted with a survey or interview by a trained assessor that proceeds with the manualized interview asking primary caretakers questions regarding communicative intent, receptive and expressive language, and language use in the home. The Vineland Adaptive Behavior Scale (Sparrow, Balla, & Cicchetti, 1984) and the MacArthur-Bates Communicative Development Inventory (Fenson et al., 1993) are typically used parent-report language assessments (e.g., Boyd et al., 2014; Fletcher-Watson & McConachie, 2017; Weismer, et al., 2010). Although parent-reports can provide an overview of the child's expressive and receptive language skills in their natural environment (Tager-Flusberg et al., 2009), parent reports may not be objective measures as parents are inclined to over-estimate their children's language skills, especially receptive language (Luyster et al., 2008). Moreover, a potential communication barrier between the interviewer and the parent or caretaker due to language or cultural background could affect the validity of the assessment (van Widenfelt, Treffers, de Beurs, Siebelink, & Koudijs, 2005).

Despite frequent use of direct assessments and parent reports to measure language skill in children with ASD, there are limitations for each approach (Luyster et al., 2008). Thus recently, more researchers recommend assessing language skills of children with ASD through language samples collected from the child with ASD in their natural setting (e.g., classroom or home;

Kasari, Brady, Lord, & Tager-Flusberg, 2013). Typically, this involves recording an audio sample from the child with ASD, transcribing the collected audio, and analyzing it according to the research question (e.g., Burgess, Audet, & Harjusola-Webb, 2013; Tager-Flusberg et al., 2009). However, collecting natural language samples can be costly in terms of resources. For example, collecting and analyzing natural language samples requires high quality and quantity of human coders. In addition, the data transfer, coding, and analysis process will take a considerable amount of time.

The Language Environment Analysis System

Current advances in technology may provide a solution for some of the difficulties of collecting language samples from children with ASD in a natural setting. The Language Environment Analysis (LENA[®]) system is one data collection system that provides an efficient way to collect and analyze natural language samples from children with ASD. The LENA system was developed in 2006 to investigate language environments of young children (Gilkerson & Richards, 2009). It consists of the LENA digital language processor (i.e., the audio recording device; DLP) that collects and records audio data, and the LENA software (i.e., the analysis software) that aggregates and analyzes the DLP-collected data (for more details, see <https://www.lena.org>).

One benefit of using the LENA system, is that the LENA software automatically produces frequency and duration data of child vocalizations, child conversations with an adult, and other environmental sounds such as adult talk, other child vocalization, and sounds from electronic devices (e.g., television). This automated process allows researchers to collect reliable and precise data in small increments for durations and repetitions that may be impossible for human data collectors. The automatic data output also eliminates preliminary human coding

needed to analyze natural language data. Another benefit of using the LENA system is the possible reduction in participant reactivity. The LENA DLP weighs less than 5.67 g and measures at 8.13 cm x 5.08 cm x 1.27 cm. The weight and size of the LENA DLP allows researchers to enclose the device in the pocket of the LENA t-shirt – a generic, round-neck t-shirt with two snap-buttons (for more information, see <https://www.lena.org>). Thus, the LENA DLP can be worn by the participants with minimal interference with the participant's daily activities ensuring stable and reliable collection of natural language data. Therefore, many researchers have been using the LENA system to investigate language development and characteristics of children with ASD using natural language data (e.g., Bak, Plavnick, & Byrne, 2019; Dykstra et al., 2012; Rankine et al., 2017; Sanders et al., 2016; Warren et al., 2010; Yoder, Oller, Richards, Gray, & Gilkerson, 2013).

Reliability of the LENA system. Although researchers have used the LENA system for language studies across a diverse sampling of children (e.g., VanDam et al., 2015 for children with hearing loss; Caskey, Stephens, Tucker, & Vohr, 2014 for infants and adults; Suskind et al., 2016 for young children; Weisleder & Fernald, 2013 for Spanish-speaking children), there has been little research devoted specifically to assess the reliability of the LENA system for children with ASD. The LENA research foundation conducted a reliability study on the LENA system-produced data for typically developing children by calculating κ coefficients between the LENA system data and human secondary raters (Gilkerson, Coulter, & Richards, 2008). However, language characteristics among children with ASD includes the presence of echolalic and stereotypic vocalizations among children with ASD with a wide spectrum of language delays, and manifests differently from peers with developmental delays and typically developing peers (Kwok, Brown, Smyth, & Cardy, 2014; Weismer et al., 2010).

One potential problem of using the LENA system in language studies for children with ASD with moderate to severe expressive language delays is that the LENA software may have difficulty accurately segmenting these prelinguistic vocalizations. Moreover, a child with ASD may communicate with prelinguistic expressive language such as grunting and babbling (Sterponi, de Kirby, & Shankey, 2015). This expressive language may be acknowledged as vegetative (e.g., coughing; LENA Research Foundation, 2015) or fixed signals (e.g., crying) by the LENA software, thus excluding prelinguistic vocalizations that are used as expressive language from the vocalization count data. In addition, although a child with ASD may demonstrate language within the appropriate developmental language age for the LENA system (i.e., between 12 to 48 months in for typically developing children), the child with ASD may exceed a chronological age of 48 months. As the LENA software separates child vocalization from adult vocalization using algorithms based on vocal frequency levels (Xu, Yapanel, Gray, 2009), the vocalization of a child over 4 years of age could be counted as adult words.

Consequently, researchers have included some reliability analysis in studies that used the LENA system for children with ASD (e.g., Bak et al., 2019; Rankine et al., 2017). In 2017, Rankine and colleagues investigated the reliability of LENA's child vocalization measure on minimally verbal children with ASD. Using Cohen's κ , Rankine et al. (2017) compared LENA collected measures of key child vocalization segments to human data collectors' measures in 56 hours of recording collected across 18 participants between the ages of 30 to 172 months. The results showed moderate reliability of the LENA system-generated vocalization data compared to human coders ($\kappa = 0.45$, $p < 0.001$) but the reliability was highly variable across participants, with reliability being higher for children who were of younger age. The researchers offered the lack of young participants (i.e., only two participants were under 48 months of age) as a possible

limitation to their study, and that this may have lowered the reliability of the LENA system in their investigation. Rankine and colleagues also pointed out that for children with ASD with severe echolalia and stereotyped language, child vocalization measures collected by the LENA system may be over-estimated due to prelinguistic language.

Similarly, Bak and colleagues (2019) calculated the intraclass correlation coefficient (ICC) between the LENA system's and human coders' child vocalization and conversational turn counts in 18 hours of recordings collected across nine participants between the ages of 6 to 10 years. Bak and colleagues (2019) found that although the LENA system's child vocalization data was reliable (ICC = .87), the reliability of the conversational turn measure was only fair (ICC = .56). The researchers suggested the possibility that the LENA system counted adults talking amongst themselves and non-related child vocalizations that occurred within the LENA DLP's range as conversational turns due to the high adult-to-student ratio in the classrooms (Bak et al., 2019). Similar to Rankine et al. (2017), Bak and colleagues also discussed the possibility that for nine minimally verbal children with ASD, conversational turn counts may be over-estimated due to stereotyped vocalizations that were not counted as conversational turns by human coders. Bak and colleagues also offered the small number of participants and samples used in their reliability analysis as a potential limitation.

The Current Study

Although previous research offers preliminary reliability analyses on LENA system measures for children with ASD, many other studies involving the use of LENA system for language studies in children with ASD did not include similar measures of reliability (e.g., Dykstra et al, 2012; Sanders et al., 2016; Warren et al., 2010). Furthermore, researchers have used the LENA system to conduct language studies in elementary schools (i.e., typically

developing children over 48 months of age; e.g., Vohr, Topol, Watson, St. Pierre, & Tucker, 2014; Wang, Miller, & Cortina, 2013), and the LENA foundation is preparing a cloud-server LENA system for practitioners. A comprehensive reliability analysis with a larger sample size than previous studies (e.g., Bak et al., 2019), various language skills, and varied age groups could substantially support the use of LENA system in research and practice for children with ASD. Thus, additional research is needed to assess whether the LENA system is reliable when used with children with ASD. Therefore, the current study investigated the reliability of child vocalization and conversational turn counts by comparing the LENA system counts and human coder counts for a heterogeneous (i.e., in terms of age, severity, sex) sample of children with ASD. Specifically, we examined the ICC between human observers and the LENA system's child vocalization and conversational turn counts for children with ASD.

Method

Participants. The current study involves a secondary analysis of language samples from 40 participants collected with the LENA system who were enrolled in year-long research examining language growth. Table 1 depicts autism severity, age in months, sex, and enrolled school/center for all participants.

School-aged participants. Language samples from 11 elementary school participants involved in a separate study investigating the effects of public-school special education on verbal language rates across the school year were included in the current study. The participants were between the ages of 6 to 9 years and were recruited from two elementary schools. Both schools had one self-contained ASD classroom each, taught by graduate-level special education teachers who had at least two years of experience teaching children with ASD. All participants from the elementary schools spent at least 77% of their time in the self-contained ASD classroom in their

respective elementary schools and the rest of their school day in grade-level general education classrooms with typically developing peers. The participants were all male and had a prior diagnosis of ASD before entering elementary school as reported on their individualized education plan.

Early childhood participants. Language samples were collected from 29 participants (seven females) between 30 to 60 months of age and enrolled in three early intensive behavioral intervention (EIBI) centers for young children with ASD. Early childhood participants were recruited from three separate EIBI centers. All centers were under management of a Midwestern University and were each housed within different local public preschools for typically developing children. All three EIBI centers delivered therapy based on principles of applied behavior analysis (ABA). Early childhood participants spent half of their day in a self-contained ABA therapy room and the other half with typically developing peers in age-equivalent preschool classrooms. All participants received one-on-one instruction and support from a behavior technician throughout the day. Early childhood participants were part of an exploratory study that investigated rates of verbal language across varied therapeutic and educational settings in the EIBI centers. The participants received an ASD diagnosis from external psychologists using the Autism Diagnostic Observation Schedule, Second Edition (Lord et al., 2012) and Autism Diagnostic Interview, Revised (Le Couteur, Lord, & Rutter, 2003) ratings prior to their enrollment in the EIBI centers.

Table 3. Age, MSEL Developmental Quotient, Location, and Sex of Participants

Age in months	MSEL DQ	location	sex
60	-	school	male
53	-	school	male
96	-	school	male
96	-	school	male
110	-	school	male
90	-	school	male
90	-	school	male
72	-	school	male
126	-	school	male
87	-	school	male
71	-	school	male
50	65	EIBI	female
57	56	EIBI	female
41	47	EIBI	female
45	65	EIBI	female
59	40	EIBI	female
52	41	EIBI	female
51	47	EIBI	female
50	72	EIBI	male
59	50	EIBI	male
55	56	EIBI	male
44	65	EIBI	male
46	70	EIBI	male
51	69	EIBI	male
33	35	EIBI	male
55	27	EIBI	male
42	34	EIBI	male
48	37	EIBI	male
41	72	EIBI	male
40	40	EIBI	male
67	26	EIBI	male
56	33	EIBI	male
50	65	EIBI	male
46	65	EIBI	male
43	21	EIBI	male
49	99	EIBI	male
40	79	EIBI	male
45	47	EIBI	male
61	48	EIBI	male
43	49	EIBI	male

Note. MSEL DQ = Mullen Scales of Early Learning Developmental Quotient (Mullen, 1995). MSEL DQ not available for participants recruited from school. Exact location of school (i.e., which of the two) or EIBI site (i.e., which of the three) not provided to protect participant identity.

Apparatus. Audio recordings of all participants were collected using the LENA DLP. The LENA DLP is a small (8.13 cm x 5.08 cm x 1.27 cm), plastic, rectangular device that can record up to 16 hr of audio. The LENA DLP is designed to fit into a pocket located in the chest area of a specially designed LENA t-shirt. When connected to a computer with LENA software, users can obtain a *comma-separated value* file detailing child vocalizations and conversational turn counts of the focal child (i.e., the child wearing the LENA DLP) in 5-min increments.

Data collection. The language sample collection was conducted at all sites (i.e., the two elementary schools and the three EIBI preschools) at least once a month throughout their time of enrollment in studies. All audio samples were collected from approximately 9:00 a.m. to approximately 3:30 p.m., except for one elementary school (with five participants) where audio samples were collected from approximately 10:00 a.m. until approximately 3:30 p.m. as requested by the classroom teacher. The primary investigator turned on the LENA DLP prior to inserting it into the LENA t-shirt and then placed the t-shirt on a participant. At the end of the school day, the classroom staff and the primary investigator took the LENA t-shirts off the participants and immediately turned off the LENA DLP.

Data extraction. For the current study, the primary investigator selected two school-day recordings from the 40 participants. The rationale for choosing two recordings was to create language samples for each participant to minimize external variables on a given day that could influence participant language patterns (e.g., illness, poor sleep). The primary investigator first extracted the language sample data from all collection days for a single participant using the LENA software. The LENA software disaggregates child vocalization and conversational turn frequencies in 5-min intervals for each day of data collection.

Fourteen of the 40 participants had only two school-day recordings, however, the

remaining 27 participants had more than three school-day recordings. To select two recordings from the 27 participants who had more than two recording days, the primary investigator employed exclusion criteria similar to Rankine et al. (2017). The primary investigator first excluded all samples with partial recordings (e.g., participant left school early, school was on a half-day schedule). Next, the primary investigator conducted outlier analyses for each collection day using a box plot and eliminated any samples that contained extreme outliers (i.e., data points further than three standard deviations from the mean) for both vocalization and conversational turn counts. If after this exclusion process, a participant still had more than two days of recordings, the primary investigator selected days that had the longest recording duration.

Human coding. The primary investigator trained four research assistants to count the child vocalization and conversational turns according to the LENA definitions (see LENA Measures, below) from the selected recordings. The research assistants had previous training in coding audio recordings and were blind to the purpose of the current study. The primary investigator selected twelve 5-min audio samples collected from a single child with ASD for training purposes. The training session began with the research assistants and the primary investigator each independently and separately counting the child vocalization and the conversational turns by listening to each audio sample. Next, the primary investigator calculated the ICC between the primary investigator and the four research assistants. The training session continued for two sessions until the primary investigator and the human coders' ICC exceeded .90 (i.e., excellent agreement; Cicchetti, 1994). The primary investigator developed data sheets to record the vocalization and the conversational turn counts for the human coders. The human coders used a standard analog counter to count and record the data while listening to the audio data.

Measures. The LENA software scores a single vocalization when the vocalization is separated from a subsequent vocalization with a pause of more than 300 milliseconds excluding fixed signals and vegetative sounds (LENA Research Foundation, 2015). A conversational turn is counted when either the focal child or an adult initiates a vocalization, the other individual (i.e., the adult or the focal child dependent on who initiated the conversation) responds within 5 s, and the first initiator responds to the response within another 5 s (LENA Research Foundation, 2015).

Data analysis. To answer the research questions - what is the reliability of the LENA system's child vocalization and conversational turn count in children with ASD measured by the ICC between human and LENA system counts – the primary investigator prepared a *Microsoft Excel* spreadsheet depicting the vocalization and conversational turn counts for each participant in 5-min increments. Next, the primary investigator randomly selected 33% of the 5-min samples using random selection software (<http://stattrek.com/statistics/random-number-generator.aspx>) from each participant. The total duration of selected recordings was 155 hr (or 1860 5-min sound files) from a total of 485.17 hr of recording (or 5822 5-min sound files). Then, the primary investigator extracted *wav*-format sound files of the selected 5-min samples with the LENA software. Each selected 5-min audio file was coded by one of the four human coders.

After all audio recordings were coded, the primary investigator entered the human coder counts in a separate Excel spreadsheet. The data (i.e., the vocalization counts and the conversational turn counts) from the LENA system and the human coders were then transferred to an *SPSS* database for statistical analyses. Using *SPSS*, the primary investigator computed the ICC between the LENA system and the human coders for the child vocalizations and conversational turn counts.

Inter-rater reliability. The primary investigator selected 25% of the 5-min audio

samples that were coded for each participant to ensure inter-rater reliability (IRR) of the human coders. The primary investigator randomly selected 466 five-minute audio samples (or 38.83 hr) for IRR. The audio samples were assigned to a human coder who did not code the file during the first round. The primary investigator derived IRR between the first and the second human coder using ICC. The mean ICC between the first and the second human coder was .97 for vocalization and .93 for conversational turn counts.

Results

The current study investigated the reliability of LENA-measured child vocalization and conversational turn count for children with ASD by examining the ICC between human coders and the LENA output with audio samples collected from 40 participants between the ages of 30 to 126 months. The mean ICC between human coders and the LENA was .76 or excellent reliability (Cicchetti, 1994) for vocalizations and .64 or good reliability (Cicchetti, 1994) for conversation turn counts. Table 4 shows the mean, range, and standard deviation of the ICC values between human coders and the LENA system for vocalization and conversational turn counts.

Table 4. Mean, range, and *SD* of ICC between human coders and the LENA system for vocalization and conversational turn counts

	<i>M</i>	<i>SD</i>	Min	Max
Vocalization	.76	.14	.33	.98
Conversation Turn	.64	.21	.00	.97

Note. MSEL DQ = Mullen Scores of Early Learning Development Quotient (Mullen, 1995)

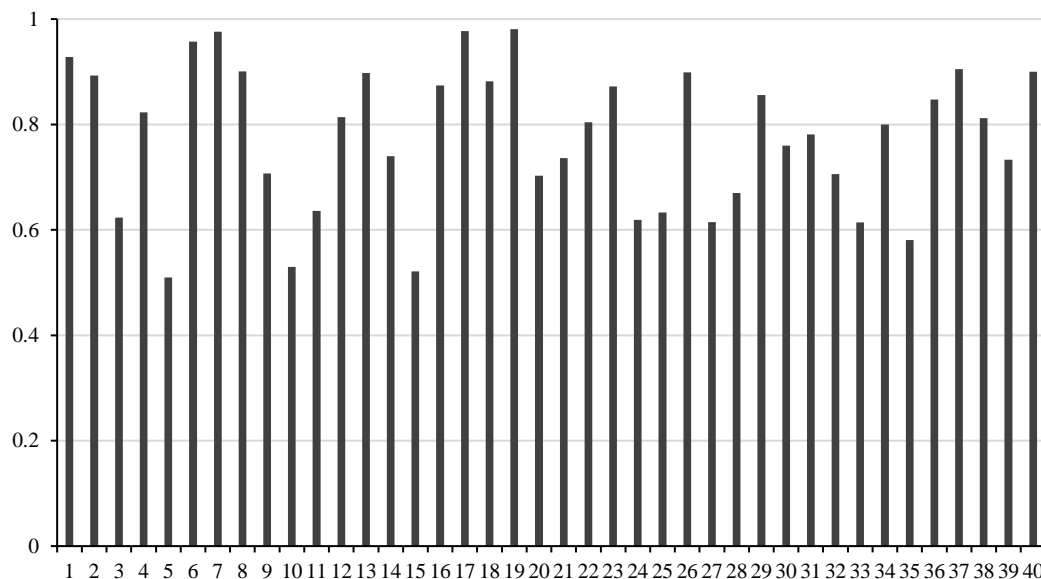
[†] Where the participants were enrolled in (e.g., School or EIBI) during data collection

* significant

Figure 4 shows that although the mean ICC can be interpreted as excellent (.76; Cicchetti, 1994), the ICC between the LENA and the human coders for 17 participants' vocalization were either good or fair (i.e., between .4 and .75; Cicchetti, 1994). Similarly, Figure 5 shows a detailed

bar graph of each participant's ICC for conversational turn counts. The *SD* for mean ICC of conversational turn counts was higher than the mean ICC of vocalizations at .21 (see Table 2). As such, there is more variation between the ICC for different participants. Although mean ICC of conversational turn counts showed good reliability (.64; Cichetti, 1994), six participants' ICC between LENA and the human coders were below .40, indicating poor reliability (Cichetti, 1994). In addition, the minimum ICC was close to zero at .004 whilst the maximum ICC was close to one at .97.

Figure 4. Detailed Bar Graph of Each Participant's ICC for Child Vocalizations

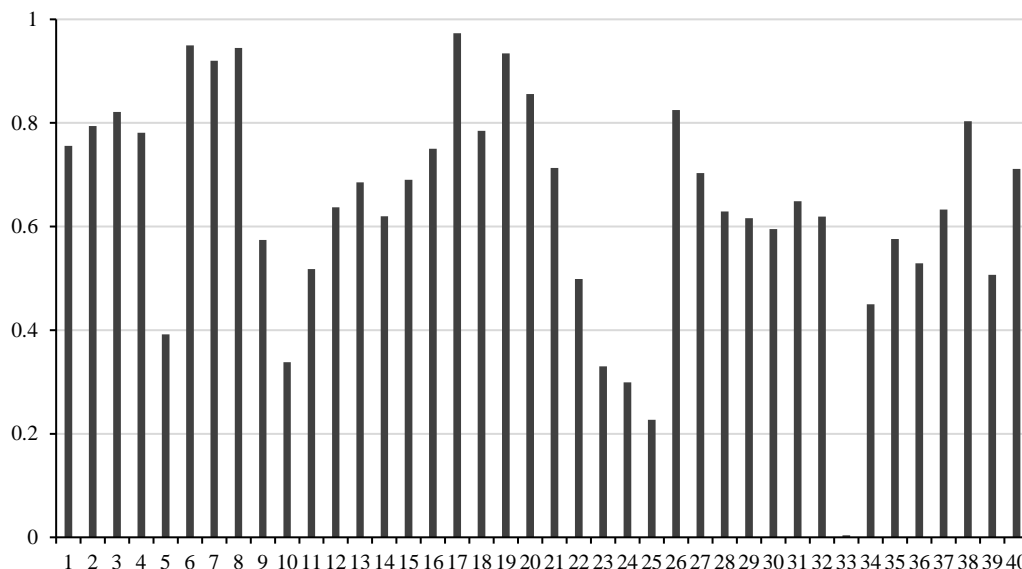


Each bar of the bar graph represents the ICC between the LENA system and the human coders for child vocalization counts. Each number on the x-axis represent each child participant and the y-axis values are ICC values.

Post-hoc analysis. The primary investigator conducted post-hoc correlational analyses to investigate any correlations between the participant's individual ICCs and 1) the age of the participants; 2) their sex; 3) ASD severity as measured by the MSEL developmental quotient; 4) the location of data collection (i.e., school or EIBI); and 5) the human coders. A correlational analysis between the participants' age in months and the vocalization and conversational turn count ICCs was not statistically significant. Sex and severity of ASD (i.e., MSEL DQ) were also

not correlated to both the vocalization and conversation turn count ICCs. However, correlational analysis revealed that the location of data collection was highly correlated ($\eta^2 = .16$; Olejnik & Algina, 2000) for conversational turns, although the correlation was not statistically significant for vocalization ICCs. In addition, human coders were strongly correlated (Olejnik & Algina, 2000) with both vocalization ($\eta^2 = .25$) and conversation turn count ($\eta^2 = .25$) ICCs. Table 5 shows the type of correlational analysis used for the post-hoc analysis depending on the type of variables and their respective results.

Figure 5. Detailed Bar Graph of Each Participant's ICC for Conversational Turns



Each bar of the bar graph represents the ICC between the LENA system and the human coders for conversational turn counts. Each number on the x-axis represent each child participant and the y-axis values are ICC values.

Table 5. Correlation analysis method and results between ICCs and variables

	Raters	Location	Sex	MSEL DQ	Age
Vocalization	.25	.16	.03	.36	.69
Conversation Turn	.25	.01	.04	.33	.81
Correlation method	eta	eta	eta	Pearson	Pearson

Note. MSEL DQ = Mullen Scores of Early Learning Development Quotient (Mullen, 1995). Raters, Location, and Sex are η^2 -values. MSEL DQ and Age are p -values.

Discussion

The objective of the current study was to provide a reliability analysis of the LENA device for use in a heterogeneous group of children with ASD to provide support for the device's use in language studies regarding children with ASD. To provide a measure of reliability, the primary investigator calculated ICCs between human coders who listened to actual audio files and the LENA system's automatic output for vocalization counts and conversation turn counts of 40 children with ASD. Results showed LENA vocalization counts may be reliable for use in heterogeneous samples of children with ASD. Although the mean ICC of conversational turn counts also showed good reliability, LENA conversation turn counts may not be reliable for some children with ASD as there was high variation in the ICCs between participants.

LENA child vocalization counts. Typical language assessments for children with ASD are often conducted during a single day with unfamiliar assessors (e.g., independent psychologists) in an unfamiliar setting (e.g., quiet room in school or clinic) resulting in measures that may not be representative of the child's skill level (Luyster et al., 2008; Rankine et al., 2017; Sandbank & Yoder, 2014). In order to provide a more comprehensive and accurate gauge of the child's language skills, researchers advise collecting natural language samples (Kasari et al., 2013). The LENA system can help researchers and practitioners who wish to collect and analyze natural, continuous and/or repeated samples directly from the children's natural environment. The LENA collected vocalization counts can supplement standardized assessments by adding a more realistic measure of the child's vocalization in the natural environment that was collected continuously over a time period. The child vocalization measure may also be used to validate certain standardized assessment results. For example, for a child who did not respond vocally to an unfamiliar assessor to a standardized assessment and scored a "floor score" as a result, the

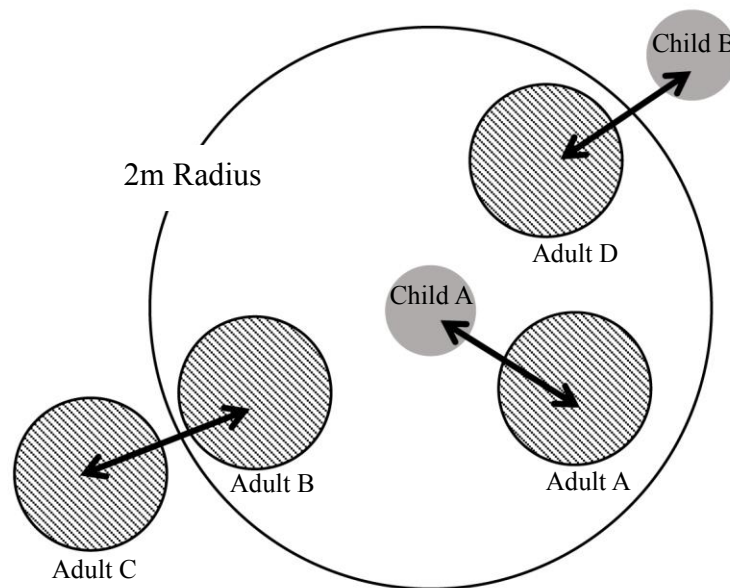
LENA vocalization counts can be used as evidence that the child's assessment score may not be an accurate depiction of the child's language skills.

Moreover, the results also showed that age or severity of ASD (as measured by the MSEL developmental quotient) were not correlated with the ICCs between human and LENA measurements. Researchers that used the LENA system concerning children with ASD above the age of four with moderate to severe ASD conducted additional reliability analysis to validate the use of the LENA system because the LENA system was originally calibrated for typically developing children between 12 to 48 months of age (e.g., Bak et al., 2019; Rankine et al., 2017). The results of the post-hoc correlational analysis between age and severity of ASD with the ICCs may be able to provide support in using the LENA system for some children with moderate to severe ASD or those who are between Grades K-3.

LENA conversational turn counts. Although the ICC for conversational turn counts between human coders and the LENA system can be interpreted as having good reliability (Cicchetti, 1994), researchers and practitioners should practice caution due to the high variability of the ICC between participants. One explanation for high variability between participants could be the difference between the LENA system and the human coders' definition of "conversational turns". As stated earlier, the LENA system uses time measures (i.e., replying with 5 s; LENA Research Foundation, 2015) whereas human coders might have been listening for context. As the human coders were provided with actual sound files, the human coders might have inadvertently included or excluded conversational turns based on the context of the conversation. Conversely, the LENA system may have counted independent self-vocalizations from the children with ASD within proximity (both time and location) of adults as conversational turns and did not count certain conversations between adults and children with ASD because the child may have taken

longer than 5 s to reply. In addition to the time parameter the LENA system uses to identify the conversational turn measure, the LENA system also uses a distance parameter. The adults within a 2-m radius of the child who interacts with the child wearing the LENA DLP is identified as conversing with the child (LENA Research Foundation, 2015). However, if an adult conversing with another adult or another adult conversing with another child is within the 2-m radius of the focal child, the LENA system may count their expressive language as conversational turns with the focal child as long as they are within the 5-s parameter (see Figure 6).

Figure 6. Conversational Turn Count Parameters of the LENA System



The gray circles represent children and the black-stripe circles represents adults. In this figure, the words of Adult B conversing with Adult C and the words of Adult D conversing with Child B may be counted as conversational turn counts with Child A if the adult words are within a 5-second parameter of Child A's child vocalization. However, the only adult truly conversing with Child A is Adult A.

Furthermore, post-hoc correlational analysis suggests that the location of the recording or the school/site the participant was enrolled in at the time of data collection was significantly correlated to the conversational turn ICCs. Additional logistic regression was conducted to determine if the difference between a public-school special education classroom and an EIBI

environment with dedicated 1:1 behavior technician could be the cause. However, there were no statistically significant findings ($p = .66$). We can only assume that in some locations, the structure of the environment may cause the LENA system to miscount certain outside conversations (e.g., adults talking to one another within close proximity of a focal child) as conversations with the focal child.

Limitations and future research. Although the current study fills a gap in the current literature for reliability analysis of the LENA system for use in a heterogeneous group of children with ASD, the current analysis has some limitations. First, although the sample includes children with ASD of different age, severity, and sex, the power of the analysis is still small to definitively conclude that the LENA system is always reliable to use in language studies with children with ASD. Future studies should recruit more participants and more recordings. Second, both the ICCs for vocalizations and conversational turn counts were significantly correlated with raters. Although the current study provided IRR between raters, the IRR was only between one second coder and the other three coders. Future studies should use multiple raters for each recording or randomly assign coders between participants, and provide an IRR between the human coders for all coders involved in the coding (e.g., IRR between Coder 1 and Coder 2; Coder 2 and Coder 3; Coder 3 and Coder 4; Coder 2 and Coder 4; Coder 1 and Coder 3; and Coder 1 and Coder 4). This may reduce possible rater effects in coding or provide additional information on the nature of the disagreement between the LENA system and human coders.

Conclusion. The current study provides some support in using the LENA system for language studies regarding children with ASD. Although the mean ICCs between human coders and the LENA system showed reliability, researchers may need to exercise caution in using certain measures (i.e., conversational turn counts). Researchers are increasingly using other

LENA provided measures such as the speech/non-speech duration (Rankine et al., 2017; Trembath et al., 2019) to supplement the LENA vocalization measure rather than the comparatively unreliable conversation turn counts. In addition, the reliability between the human coder and the LENA system suggest that the LENA recordings may also be used for human coding as an audio recording device for qualitative studies regarding the context and the quality of communication between adults and children with ASD (Burgess et al., 2013; Tager-Flusberg et al., 2009).

Language is a pivotal skill that can provide overall development in children with ASD (Pickles et al., 2014). As such, many researchers continue to study language interventions and language development in children with ASD to inform education and research (Sandbank et al., 2017). Natural language samples can provide in-depth understanding about a child's language skill levels when used separately or in conjunction with a standardized assessment (Kasari et al., 2013; Luyster et al., 2008; van Widenfelt et al., 2005). The LENA system can be an efficient method to collect continuous/repeated language samples in the natural environment for children with ASD for various language research.

CHAPTER 4

Original Study

Language and Children with ASD

Autism spectrum disorder (ASD) is a pervasive developmental disorder that affects approximately one in 40 children in the United States (Kogen et al., 2018). Symptoms and severity vary across children, but individuals with ASD commonly experience difficulties in language and social communication (American Psychiatric Association, 2013). Language development trajectories of children with ASD during infancy and toddlerhood is comparatively lower than that of typically developing children and children with other developmental disabilities (Gernsbacher, Morson, & Grace, 2016). And unlike typically developing children, young children with ASD may not develop expressive and receptive language repertoires through naturally occurring social interactions (Fisher & Meyer, 2002; Whitaker, 2004). Therefore, it is important to investigate and understand elements such as interventions and language environments that can maximize language development for children with ASD.

Language and communication skills are critical factors for an independent adult life (Roux, Shattuck, Rast, Rava, & Anderson, 2015). According to the National Autism Indicators Report (Roux et al., 2015), young adults with ASD with minimal or no language skills are 5 times more likely to have no interaction with a peer after graduating high school and 10 times more likely not to have higher education or jobs in their transition plans compared to other young adults with ASD. Furthermore, children with ASD who do not demonstrate functional communication skills by age 6 may remain nonverbal throughout life (Tager-Flusberg & Kasari, 2013). Early intervention can provide a promising opportunity to overcome language deficits and demonstrate developmental gains in language for children with ASD (Eigsti, de Marchena,

Schuh, & Kelley, 2011). The National Research Council (2001) reported that the earlier interventions are implemented, the better the language developmental outcomes are for children with ASD. Consequently, there are many interventions that focus on increasing expressive language and social communication in young children with ASD.

Language Interventions

Language interventions for young children with ASD can be considered as either focused or comprehensive intervention models. Focused interventions concentrate on teaching specific language units (Kane, Connell, & Pellicchia, 2010; Virués-Ortega, 2010). Focused interventions typically target specific language units such as requesting (e.g., asking for assistance; Reichle, Dropik, Alden-Anderson, & Haley, 2008), identifying grade-level vocabulary (e.g., color identification; Akande, 2000), or social interactions (e.g., greetings; Reichow & Sabornie, 2009). For focused interventions, the intervention's effect is documented by the increase in accuracy or use of the targeted language unit over time (e.g., Greer & Ross, 2008). Researchers have identified many focused interventions as evidence-based practices for teaching language skills to children with ASD (see Wong et al., 2015). But because focused interventions measure the occurrence of a specific language unit, it is difficult to gauge the effects of the focused intervention on overall language development (Norrelgen et al., 2015).

Conversely, comprehensive interventions typically consist of a packaged curriculum of evidence-based practices that target the overall development in young children with ASD including language, social, and functional skills (Odom, Boyd, Hall, & Hume, 2010). Hence, the intervention's effect on language is documented through standardized pre- and post-assessments (e.g., *Preschool Language Scale*; Zimmerman, Steiner, & Pond, 2002) or researcher-created language probes (e.g., observation of language; Tager-Flusberg et al., 2009). Studies have shown

that some comprehensive interventions have a positive effect on the language development in some children with ASD (e.g., Boyd et al., 2014; Makrygianni & Reed, 2010). However, aggregated mean-score results can be difficult to apply to children with severe symptoms of ASD who, despite receiving comprehensive language interventions, may show little to no improvement in language and are often treated as outliers in statistical analyses (Bak, Plavnick, & Byrne, 2019; IACC, 2014).

Despite limitations, both focused and comprehensive interventions show environmental change can facilitate an increase in expressive language and may positively affect language development among children with ASD. Researchers have discussed the possibility of certain environmental variables that may facilitate the interventions within a child's learning environment such as a school or the home (Boyd et al., 2014; Kane et al., 2010). These environmental variables include where the intervention was implemented, who implemented the intervention, and how many other children were also included in the intervention setting. Identifying environmental variables outside the intervention that can promote the effectiveness of an intervention and manipulating those variables can increase the potency and permeation of an intervention (Boyd et al., 2014).

In 2014, Boyd and colleagues investigated the effects of comprehensive interventions for 198 children with ASD. Boyd et al. (2014) compared three types of classrooms that implemented the LEAP (Learning Experiences and Alternative Program for Preschoolers and their Parents) program, the TEACCH® Autism program, and a mix of eclectic ASD interventions typically conducted in a “high-quality” public special education classroom. Boyd and colleagues (2014) found that all participants in the three classrooms had significant gains in communication and fine motor skills after 6 months of treatment. This result differed from previous studies (e.g.,

Baril & Humphreys, 2017; Makrygianni & Reed, 2010; Peters-Scheffer, Didden, Korzilius, & Sturmey, 2011) that found large gains in the treatment group (i.e., classrooms that implemented a comprehensive intervention) and little to no gains in the control group (i.e., no specific comprehensive interventions). Boyd and colleagues (2014) suggested that outside environmental variables such as classroom management and teacher interaction rather than differences in each intervention may explain why the three different interventions resulted in similar effectiveness for the young children with ASD in their study.

Measuring Environmental Effects on Language

Identifying environmental variables that positively effect expressive language in children with ASD or supplement the effectiveness of interventions may further facilitate language development in children with ASD. Additionally, practitioners may be able to provide optimal teaching opportunities for the children with ASD. To identify possible environmental effects, researchers will need to simultaneously collect children's language and the environmental variables in the children's natural learning environment consistently throughout the course of the study. This will require increased precision, scope, and scale of the data collection and may subsequently increase logistical and financial costs such as human resources and time. However, technological advances offer automated language data collection systems, such as the *Language Environment Analysis* (LENA[®]) system. The LENA system records natural language and analyzes collected audio recordings to provide frequency and duration data. Using automated data collection apparatuses such as the LENA system can allow researchers to collect language and environment data automatically and reduce costs and time necessary for manual data collection.

In 2010, Warren and colleagues (2010) first used the LENA system with 26 young

children with ASD to study their home language environment in comparison to 78 typically developing children. Warren and colleagues (2010) found that although the difference in the amount of adult talk was not statistically significant between households of typically developing children and children with ASD, the frequency of conversation and child vocalizations was significantly lower in children with ASD. The researchers stated that although parents of children with ASD try to foster a “rich language learning environment”, the conversations between the children with ASD and their parents did not lead to prolonged communicative interaction as did the typically developing children and their parents (Warren et al., 2010, p.569).

Extending the study by Warren and colleagues (2010), Dykstra et al., (2012) used the LENA system to investigate preschool language environments for 40 young children with ASD. Dykstra and colleagues (2012) found that participants with ASD who had poorer language and cognitive skills were less likely to be talked to by an adult compared to their peers with higher language abilities (Dykstra et al., 2012). The researchers discussed the importance of a “high quality” early education environment for language development and suggested future research that use the LENA system to evaluate different language learning contexts and opportunities within the preschool classroom for children with ASD (Dykstra et al., 2012, p. 592).

Additionally, Burgess, Audet, and Harjusola-Webb (2013) conducted a mixed-method study to determine environmental factors such as complexity of adult language that promote increased expressive language in children with ASD. The researchers found that although there was little difference in the complexity and the amount of adult talk in both the school and the home learning environment, the school environment focused more on “adult led/directed” instructions rather than interactive social communications compared to the typically developing children (Burgess et al., 2013, p.436). Burgess and colleagues (2013) suggested that child-

focused activities such as naturalistic interventions may facilitate meaningful and sustained conversation and promote language development in the classroom.

The previous studies that investigated the language environment of children with ASD in the homes and schools largely focused on one environmental variable - adult language. Few studies have investigated other environmental variables such as location, instructional grouping, intervention delivery method (e.g., naturalistic or contrived), and their possible effects on language in children with ASD. Therefore, the current study investigated the language learning environment of an early intensive behavior intervention (EIBI) center to analyze if there are any differences in the quantity of expressive language and social communication under different environmental variables - location, grouping, activity, and type of instruction. Specifically, this study asks the following research questions:

1. Is there a difference in the quantity of selected LENA system language measures under different environmental variables?
2. Is there a correlation between selected LENA system language measures and different environmental variables?
3. And if correlations are identified, are there statistically significant differences in the quantity of selected LENA system language measures under different environmental variables?

Method

Participants and setting. Participants in this study included 21 preschool-age children with ASD and 32 behavior technicians (BTs) from three EIBI centers located in Midwestern United States. The three EIBI centers were housed within three separate community preschools. All three centers operated under the executive direction of a single university faculty member and followed identical curriculum and implementation procedures. The first center was housed

within a child development laboratory preschool located within a large Midwestern University in a university town. The second center was housed within a public Head Start preschool in an urban city. And the third center was housed within a public preschool in a suburban area.

All 21 children (three female children and 18 male) received an ASD diagnosis from an independent psychologist through a full psychological evaluation battery including the Autism Diagnostic Observation Schedule, Second Edition (Lord et al., 2012) prior to enrollment in the centers. Eight children were enrolled in the first center, seven children in the second center, and six children in the third center. The children received comprehensive interventions based on applied behavior analysis (ABA) for young children with ASD in a self-contained EIBI classroom. The children also spent some time of their day in the general education preschool classroom with same-age typically developing children according to each child's independent educational goal and skill level. Each child with ASD received one-to-one instructional and behavioral support from a BT throughout the day from 8:30 a.m. to 3:50 p.m.

The BTs included in the current study included 27 females and five males. Of the 32 BTs, 16 were enrolled in a master's program in applied behavior analysis (ABA). The BTs varied in their experiences working with young children with ASD. The centers each employed 14 BTs (two males), nine BTs (one male), and nine BTs (two males) respectively, equally employing more experienced BTs and less experienced BTs. Although some of the BTs worked a half-day shift, either from morning to lunch or lunch to afternoon, this 1:1 BT to child ratio was kept consistent throughout the day. The mean and range of BTs' ages, and the mean and range of the children's age and Mullen Scales of Early Learning (MSEL; Mullen, 1995) developmental quotients are provided in Table 6. Characteristics per participant are not provided to protect their identities.

Apparatus. The LENA system consists of the digital language processor (DLP) and the LENA software for data aggregation. The DLP is a small rectangular plastic object (8.57 cm x 5.56 cm x 1.27cm) that weighs about 50 g. The DLP fits into a pocket located in the chest area of a generic cotton t-shirt approximately 10 cm from the neck. When a child wears the t-shirt with an operating DLP, the DLP records all audio data within 2 m from the child for up to 16 hr (Gilkerson & Richards, 2009). From the DLP-collected data, the LENA software produces frequency and duration data of child language from this audio data. Reliability of the LENA system with young children with ASD is established in previous literature (see Xu, Yapanel, & Gray, 2009; also, Yoder, Oller, Richards, Gray, & Gilkerson, 2013) and in Chapter 3 of the current dissertation.

Table 6. Age Mean and Range of Adult and Child Participants

		<i>M</i>	Min	Max
BTs	Age ^a	23.94	21	29
Children	Age ^b	48	28	60
	MSEL DQ	55	33.1	78.8

Note. MSEL DQ = Mullen Scales of Early Learning Developmental Quotient (Mullen, 1995); BT = Behavior Technician.

^a Age in years on October 1, 2018.

^b Age in months on October 1, 2018.

Data.

LENA data. The current study used the child vocalization frequency, child conversational turn count, and the adult word count measure automatically aggregated by the LENA software from the DLP. Child vocalization frequency represents the number of vocalizations emitted by the child participant during the audio recording. A vocalization is counted as a single vocalization when it is separated from a subsequent vocalization by a pause greater than 300 ms (Gilkerson & Richards, 2009). The conversational turn count represents the number of three-step

interactions between a child and an adult. For example, if a child speaks, the adult answers within 5 s, and the child replies to the adult within 5 s, this three-step interaction is counted as one conversational turn. The LENA software also counts three-step interactions initiated by the adult, the child answering within 5 s, and the adult replying to the child's answer within 5 s as a conversation turn. The adult word count represents the number of words spoken by the adult within a 2-m radius of the child wearing the DLP.

Environment data. The environment data included the exact start time, the location, the delivery method, the grouping, and the objective of the activity. Location was coded into two variables – EIBI and inclusion, where inclusion was defined as locations where the child was exposed to typically developing peers. Grouping was coded into two variables – individual and group, where group meant that the child was participating in an activity with other children. Delivery method was also coded into two variables – discrete and natural, where discrete meant that the activity was systematically preplanned and regulated the boundaries of the activity for both the BT and the child (see also, Peters-Scheffer et al., 2011). Finally, objective was coded into four variables – language, adaptive, social/play, and cognitive, depending on the learning objective and target skill of the activity.

The BTs assigned to the child participants at the EIBI centers collected the environment data on a pre-developed data sheet. The primary investigator met with the three Board-Certified Behavior Analysts (BCBAs) from each EIBI center two-weeks prior to the start of the study to discuss the data sheet and environment data collection. Then each BCBA trained BTs in their respective center to collect data using the data sheet prior to the start of the study. The primary investigator also attended this training session to assure fidelity and procedural integrity. A sample of this data sheet is presented as Appendix B.

Procedure.

Data collection. The primary investigator collected the LENA audio data from each center on three school days, once a week, during different days across three consecutive weeks of a single month. Three repeated data collections for each center were conducted because a single day's audio data may not be representative of the actual language skill and pattern for children with ASD (Sandbank & Yoder, 2014). On the day of data collection, the t-shirt holding the DLPs were distributed to each child at approximately 8:30 a.m. and collected at approximately 3:40 p.m. During this time, the BTs also recorded environment data on the environmental data sheet.

All children had previous exposure to the LENA t-shirt as monthly language samples had been collected at the centers for another study and demonstrated no discomfort wearing the t-shirts with the DLPs throughout the day. The collected data from the DLPs were immediately transferred to the LENA software and the audio measures were aggregated by 5-min intervals — the shortest interval provided by the LENA software. The 5-min interval was selected to match the durations of many isolated instructional sessions administered at the EIBI centers.

Database. To build the database, first, the primary investigator aligned the BT-collected environmental data to the 5-min interval data provided by the LENA software. The primary investigator entered the codes for location (EIBI was coded as 0, inclusion was coded as 1); grouping (individual was coded as 0, group was coded as 1); delivery method (discrete was coded as 0, natural was coded as 1); and objective (language was coded as 1, adaptive was coded as 2, social/play was coded as 3, cognitive was coded as 4) for each 5-min interval data by matching the start time recorded by the BTs on the environmental data sheet.

Second, the primary investigator excluded all 5-min intervals that did not contain a *full set* of environmental variables. A 5-min interval was considered as a *full set* if there were no

missing data for the four environmental variables for location, grouping, delivery method, and objective; and if all four variables did not change throughout the LENA presented 5-min interval. For example, if the objective of the activity changed from adaptive to language at 9:03 a.m., the 5-min interval data from 9:00 a.m. to 9:05 a.m. was discarded even if other variables such as location, grouping, and delivery method remained consistent.

Finally, based on the collected measures and coded variables, the primary investigator created a database using *Microsoft Excel* where each row represented 5-min intervals of data and each column represented a variable or measure. The measures and variables included in the primary analysis are as follows:

1. Measures from the LENA system included child vocalization, conversational turns, and adult words per 5-min segment.
2. Variables from the environmental data included location, grouping, delivery method, and objective during the 5-min segment.
3. Participant information included each child's identification number, a EIBI center number, and the collection date.

Analyses. First, the primary investigator computed descriptive statistics for the LENA system derived language measures (here after referred to as the dependent variables) between and within child participants under the different environmental variables (here after referred to as the independent variables) using *SPSS*. Second, the primary investigator specified an unconditional model to check for possible nesting because the data structure was similar to school-like structures where variables can be dependent on a multi-level. For example, child-outcomes are dependent on instruction, instruction is dependent on individual teachers, and teachers can be dependent on school policy (see Raudenbush, 1988). The primary investigator

checked the intraclass correlation coefficients (ICCs) of each hypothesized nested-level (Heck, 2001) using *Stata*. However, the ICC values showed that the data were likely not nested (i.e., the ICC value was less than .05 for all variables; Heck, 2001). Third, the primary investigator conducted a correlational analysis in *SPSS* to identify correlations between the dependent variables and the independent variables. Finally, based on the correlation results, the primary investigator used *Stata* to conduct Linear Mixed Model (LMM) analyses to investigate any relationships between the dependent variables and the independent variables. An LMM analysis allows researchers to account for data dependency in repeated measures, and unbalanced data structures where more data may have been collected for certain environmental variables (Cnann, Laird, & Slasor, 1997; Raudenbush & Bryk, 2002). Prior to all analyses, the primary investigator analyzed the data and performed transformations to meet respective assumptions if needed.

Reliability. Reliability was assessed for both the independent and the dependent variables. For the independent variables, inter-rater reliability (IRR) was conducted to check if the primary investigator transferred information from the BT-collected environment data sheet to the database accurately. For this, the primary investigator trained two graduate-level independent researchers who had worked as a BT for at least 6 months at one of the EIBI centers prior to the current study for IRR. First, the primary investigator trained the independent researchers using several sets of 30 randomly selected 5-min intervals. Second, the independent researchers entered the environmental information in numbered codes (see *Database* section) for each 5-min interval by referring to the environmental data collected by BTs during the corresponding time. The training continued until each independent researcher's coded data agreed at least 90% with the primary investigator. One independent researcher met criterion of agreement after one training-set and the other independent researcher met criterion after coding three training-sets of

5-min intervals. Finally, the primary investigator randomly selected one-third of each child participant's 5-min intervals across the three collection days and randomly divided them between the two independent researchers. A total of 819 5-min intervals was selected for IRR. The final IRR for each variable was calculated in *SPSS* with Cohen's κ . The IRR between a second rater and the primary investigator was .84 for location, .87 for grouping, .77 for delivery method, and .79 for the objective.

For the dependent variables collected by the LENA system, the primary investigator conducted an independent reliability analysis to support the use of the LENA system for children with ASD. This reliability analysis is presented in Chapter 3. This reliability analysis was conducted by calculating the ICC between the LENA-system and human coders for child vocalization and conversational turn counts from 40 children with ASD. The results showed that the mean ICC between the LENA system and human coders was .76 for child vocalization and .64 for conversational turn counts for the 40 children with ASD that participated in the reliability study (see Chapter 3 for more information).

Results

Audio and environment data were collected for a total of 25,135 min (418.91 hr) for nine days spread across three weeks (one day a week for each EIBI center for three weeks) from 21 children with ASD and 32 BTs with a range of 425 min (7.08 hr) to 290 min (4.83 hr) per day. After coding for environmental data, the number of 5-min interval that contained a *full set* of the environmental data was 2445 (12225 min or 203.75 hr; 48.63% of the original data collected) from all participants across three days for each site.

Table 7. The Mean, Range, and *SD* for Child Vocalization, Conversation Turn, and Adult Word Counts for Each Environmental Variable.

			<i>M</i> [†]	min [†]	max [†]	<i>SD</i> [†]
Location	EIBI (n = 2252)	CVC	14.65	0	110	12.39
		Turn	3.74	0	34	3.97
		Adult	109.54	0	733	96.29
	Inclusion (n = 193)	CVC	19.24	0	71	14.03
		Turn	4.16	0	20	3.78
		Adult	86.52	0	448	82
Grouping	Individual (n = 943)	CVC	13.55	0	78	19.24
		Turn	3.88	0	34	4.22
		Adult	114.44	0	485	100.38
	Group (n = 1502)	CVC	15.93	0	110	14.65
		Turn	3.71	0	34	3.77
		Adult	106.99	0	733	86.52
Delivery Method	Discrete (n = 539)	CVC	13.46	0	74	13.46
		Turn	4.03	0	34	4.53
		Adult	113.16	0	471	93.48
	Natural (n = 1906)	CVC	15.45	0	110	15.45
		Turn	3.7	0	34	3.77
		Adult	103.18	0	733	95.94
Objective	Language (n = 241)	CVC	14.74	0	57	11.66
		Turn	4.56	0	34	5.13
		Adult	105.71	5	455	83.83
	Adaptive (n = 649)	CVC	13.55	0	110	12.36
		Turn	3.69	0	23	3.87
		Adult	124.35	0	733	111.2
	Social/Play (n = 1407)	CVC	15.76	0	92	12.67
		Turn	3.65	0	34	3.74
		Adult	100.63	0	665	89.89
	Cognitive (n = 148)	CVC	14.78	0	74	13.79
		Turn	4.08	0	20	4.1
		Adult	105.51	0	384	80.14

Note. CVC = child vocalization count, Turn = conversational turn count, Adult = adult word count.

[†]counts per 5-min.

Child vocalization count means were higher in inclusion locations compared to the self-contained EIBI classroom, group instruction compared to individual instruction, natural delivery compared to discrete delivery, and during social/play instruction compared to the other learning objectives coded. Conversational turn count means were also higher in inclusion settings, but

unlike child vocalization count means, they were higher during individual instruction, discrete delivery, and during language instruction. Adult word count means were higher in the self-contained EIBI location, individual instruction, discrete delivery, and during adaptive instructional objectives. Table 7 depicts the frequency, mean, range, and *SD* of child vocalization, conversational turn, and adult word counts for each coded variable.

Correlational analyses conducted between the dependent variables and the independent variables showed that child vocalization was significantly correlated to the location, grouping, and method of delivery ($p < .001$ for all); and adult word counts were significantly correlated to the location ($p < .001$). The correlational analyses results are presented in Table 8.

Table 8. Correlational Analysis Results Between Language Measures and Environmental Variables.

	Location^a	Grouping^a	Method^a	Objective^b
CVC	< .001 ^{***}	< .001 ^{***}	.001 ^{**}	0.006
Turn	.16	.26	.09	0.00
Adult	.001 ^{**}	.63	.13	0.001

Note. CVC = child vocalization count, Turn = conversational turn count, Adult = adult word count.

^a *Pearson* correlation p -values.

^b Eta² correlation η^2 values.

^{**} $p < 0.01$, ^{***} $p < 0.001$.

The primary investigator conducted LMM analysis based on the results from the correlational analysis. Before the LMM analysis, child vocalization and the adult word count data were transformed to meet the assumptions of the analysis. Child vocalization frequencies were converted to their square roots and the adult word count was converted to their log10 values. Extreme outliers that were more than three inter-quartile distances from the mean were also removed from each dataset based on descriptive statistics and box plots. For the LMM analysis, first, the primary investigator conducted LMM analysis between child vocalization frequencies and the location, grouping, and the delivery method of the activities. Child

vocalization was higher in the inclusion setting compared to the self-contained ASD classroom ($\beta = 0.55$; $p < .001$), when the children were receiving group instruction compared to individual instruction ($\beta = 0.27$; $p < .001$), and when the activity was delivered naturally ($\beta = 0.26$; $p < .001$). Next, the primary investigator conducted LMM analysis between the adult word counts and the location of the activity. Adults were more likely to talk to the children in the self-contained EIBI classroom than in inclusion locations ($\beta = 0.1$; $p < .001$). Table 9 shows the detailed results of the LMM analyses including the *SE* and the *CI*.

Table 9. Regression Coefficients from Linear Mixed Models for Child Vocalization and Adult Word Counts

		β	<i>SE</i>	<i>z</i>	95% <i>CI</i>	
CVC^c	Location	0.55***	0.11	4.97	0.34	0.77
	Group	0.27***	0.06	4.32	0.15	0.39
	Method	0.26***	0.07	3.52	0.11	0.4
Adult^d	Location	0.10***	0.03	3.55	0.05	0.16

Note. CVC = child vocalization count, Turn = conversational turn count, Adult = adult word count.

^c measures were converted to square roots.

^d measures were converted to log10 values.

*** $p < 0.001$.

Discussion

The current exploratory study aimed to identify possible environmental variables that positively affect the quantity of children with ASD and adults' expressive language as measured by the LENA system in three EIBI centers. Whereas, previous studies that used the LENA system to investigate children with ASD's expressive language focused on the effect of adult's language (e.g., Burgess et al., 2013; Dykstra et al., 2012; Warren et al., 2010), the current study analyzed environment variables such as the location, instructional grouping, delivery method, and the instructional objective. Results showed there were differences in the child vocalization counts between different environmental variables, with child vocalizations significantly higher

during inclusion compared to the self-contained EIBI classroom, during group instruction compared to individual instruction, and during interventions delivered naturally compared to discrete. The current study's result contributes to existing literature by providing preliminary evidence that child vocalization from children with ASD may differ depending on certain environmental variables.

Location. Mean child vocalization and conversational turn counts were all higher in inclusive settings even though adults on average produced more words in the self-contained EIBI classroom. In addition, the LMM statistical results showed that children with ASD displayed higher levels of vocalization in the inclusion setting compared to the self-contained EIBI classroom. Although unable to be discerned within the current scope of the study, there are multiple hypotheses that may explain the reason for the increase in child vocalization. The current results might suggest children with ASD produce more expressive language in inclusion compared to the self-contained EIBI classroom. Conversely, the results may also suggest there was an increase in prelinguistic vocalizations such as stereotypical and echoic vocalizations in children with ASD in these environments. Or the results may also indicate there were both higher instances of meaningful expressive language and prelinguistic vocalizations.

The inclusion environment may have provided more opportunities for the BTs to engage in instructional conversation with the children with ASD due to the spontaneity of an inclusive preschool classroom (see Burgess et al., 2013). This may explain why mean adult words were lower in the inclusive classroom compared to the self-contained EIBI classroom but mean conversational turns between the children with ASD and BTs were higher during inclusion. There is another possibility that the inclusion environment may have provided more opportunities for the children with ASD to increase vocalizations — both meaningful expressive

language and prelinguistic vocalizations — because of typically developing peers or unfamiliar stimuli. Whatever the reason, we cannot draw firm conclusions from the current data because the current study did not include any contextual or qualitative analyses for the collected audio recordings. There are a lot of discussions about whether the inclusion classroom or the self-contained ASD classroom is a more effective learning environment for children with ASD (e.g., Simpson, Mundschenk, & Heflin, 2011). Although the current results cannot provide a definitive answer, further investigation into the contents of the children's expressive language may provide much needed information regarding the effects of the two environments on the language of children with ASD.

Grouping. Descriptive and statistical analysis results also showed child vocalizations were significantly higher during group instruction compared to individual instruction. The results could be mean the children with ASD produced more expressive language or that there was also an increase in stereotypical and echoic vocalizations. During group instruction, children with ASD may have increased expressive language to interact with the peers in the group setting. Conversely, there may have been an increase in prelinguistic vocalizations because individually-assigned BTs provide minimal behavioral support to maintain the child's focus on the BT leading the group. And this may have caused the BTs to inadvertently miss stereotypical and echoic vocalizations that would have been redirected during individualized instruction. Additional research should be conducted to investigate the content of the vocalizations during group settings because group instruction becomes more prevalent in public school settings. If the increase in child vocalizations is due to increase in meaningful expressive language, increasing EIBI interventions in small groups or dyads may increase language development, better prepare young children with ASD for future school learning, and have additional implications for human

resource management.

Delivery method. Children with ASD also showed significantly higher probability of increased vocalizations during naturally delivered interventions compared to discretely delivered interventions. Many researchers have developed naturalistic language interventions for children with ASD to increase language opportunities for the child with ASD and for the supporting adult to offer more naturally occurring teaching opportunities (Ingersoll & Schreibman, 2006; Kaiser, Hancock, & Nietfeld, 2000). However, since a qualitative analysis was not included in the scope of this study, we do not know the content of children's vocalizations or conversations between children and adults, there is a possibility the higher child vocalization frequency could be the result of an increase in stereotypical and echoic vocalizations in children with ASD during this less restricted instruction time. If the child vocalizations were increased because of prelinguistic vocalizations, it may suggest that more discrete intervention elements may be needed for children with severe language delays or that naturalistic language interventions may not be as effective for children with ASD who display prelinguistic vocalizations. Further qualitative analysis to learn about the nature and the context of the child vocalizations is needed to inform current naturalistic language interventions.

Implications. The current study showed that there are indeed differences in child vocalizations between environments, but we do not know whether the increases point to a positive or negative language environment for children with ASD. Inclusion periods in the general education preschool classroom provide young children with ASD the opportunity to socially interact with typically developing peers and learn age-appropriate pre-academic, functional, and social skills (Simpson, De Boer-Ott, & Smith-Myles, 2003). In group instruction, common learning objectives for children with ASD are joint-attention, observational learning,

and social interaction (Dotson, Leaf, Sheldon, & Sherman, 2010). And natural interventions provide natural learning opportunities for more effective generalization in children with ASD (LeBlanc, Dillon, & Sautter, 2009). But if children with ASD produced higher levels of stereotypical and echoic vocalizations in these environments, there is a possibility that they may not be receiving the full benefits of the respective interventions (Jang, Dixon, Tarbox, & Granpeesheh, 2011). More research is needed to investigate the nature of the vocalizations produced by children with ASD under different environmental variables. Research that reexamines environmental variables in which the interventions are implemented may allow researchers to identify ways to optimize existing evidence-based interventions to increase effectiveness specifically for language development in children with ASD (Boyd et al., 2014; Kane et al., 2010). Optimizing current interventions to maximize learning opportunities for children with ASD will help them make increased gains in language development during early intervention.

Furthermore, investigating the contents of the child vocalizations will provide information on whether environmental variables effect all children with ASD in general or if there is a difference affected by ASD symptom severity. For example, children with severe ASD may have produced more prelinguistic vocalizations whereas children with moderate to mild ASD may have produced more meaningful expressive language in the same inclusion environment. Such research could enhance current knowledge about the effects of comprehensive interventions for children with severe symptoms of ASD as they are often treated as outliers or excluded from intervention assessment research (Interagency Autism Coordinating Committee, 2013).

The current chapter presents an exploratory study that focused on the quantitative

measurement of children with ASD's natural language. As mentioned in Chapter 2, automated data collection apparatuses can help researchers increase efficiency, accuracy, and scope for data collection. In the current study, the LENA system provided a way to collect continuous and repeated samples of the children's language under different environmental variables that may not have been possible within the resources available to the primary investigator. Despite the benefits of using the LENA system as an automated data collection apparatus, there were limitations to the implications of the results because the current study only disseminated quantitative data. Future research should include qualitative research involving human data collectors that can provide in-depth analysis based on the preliminary quantitative analysis provided by the automated data collection apparatus for language research regarding children with ASD. This may allow researchers to examine specific language in broader populations and environments possibly combining the strengths of assessment in focused and comprehensive language interventions.

Limitations and future research. There are several limitations to the current study. First, as mentioned, the current study did not analyze the content of the children's vocalizations or their conversations with the BTs. Future studies should incorporate transcription analyses to further investigate the relationship between the context and quality of the children's expressive language and environmental variables. Second, the current study did not implement reliability analysis for the BT-collected environmental data. In the current scope of the study, there was no method to determine to what extent the environmental variables recorded by the BTs were accurate because listening to the audio data provided by the LENA system was not sufficient to discern the specifics of environment. Future studies should implement IRRs between BTs and independent researchers either in-vivo or through video recordings to establish reliability of the environment

data. Third, we collected three repeated samples from each center to counter the possibility that a single day's data may not be representative of typical child vocalizations (Sandbank & Yoder, 2014) or interventions carried out in the EIBI centers. However, a post-hoc homoscedasticity analysis of the LENA measures by participants per collection day showed the majority of the participants' data did not show any statistical difference between the three collection days. Thus, the repeated collection and the subsequent inflation of the number of samples may have inadvertently inflated statistical significance in some of the analyses. In addition, because the current study only included data that were considered *full-sets*, only 48.63% of the collected data was included in the analysis. Therefore, the statistical results may not accurately represent the entire data collected. Finally, researchers should take caution in generalizing the results of this study to other settings and other children with ASD. The current study was conducted in a specific EIBI environment with a small sample size. More studies with larger samples including children with ASD receiving different types of early intervention such as a public school early special education program and children with moderate to severe ASD that are often left out of research (IACC, 2014) should be conducted to gain more information regarding environmental variables that can positively affect language interventions for young children with ASD.

Conclusion. Language is an important skill for children with ASD to successfully enter society as independent adults (Roux et al., 2015). Although early intervention is crucial in helping young children with ASD develop language skills (Eigsti et al., 2011), identifying environment variables that strengthen existing evidence-based interventions may be an efficient way to increase the potency of intervention and further facilitate long-term language development. The present investigation offers an approach to collecting language and environment data that could be replicated and expanded upon to better understand what types of

vocalizations occur in the varied instructional environments. Additional research may offer potential variables for manipulation such as the delivery of the intervention or the grouping of children in a more targeted and specific intervention-based research that may increase effectiveness in language interventions.

CHAPTER 5

Discussion

The current dissertation investigated the impact of language environments on the language of children with autism spectrum disorder (ASD) using automated data collection technology combined with direct observation by researchers. Chapter 2 was a systematic literature review on the current state of applied behavior analysis (ABA) research that used automated data collection apparatuses. The results showed that only 10.16% of 1466 data-base studies collected from five ABA journals between 2010 and 2018 used an automated data collection apparatus. This supports Crowley-Koch and Van Houten's (2013) claim that the field of ABA is not utilizing technological advancements that can benefit observation and measurement of human behavior.

Chapter 3 presented a reliability study of the Language Environment Analysis (LENA) system – an automated data collection apparatus for language data. Intraclass correlation coefficient (ICC) values between human coders and the LENA system were used to assess the reliability of the LENA system measures for 40 children with ASD. The results showed the ICC between humans and the LENA system were .76 for child vocalization counts and .64 for conversational turn counts. The results suggest that LENA child vocalization measures may be highly reliable (Cicchetti, 1994). However, researchers should exercise caution when using conversational turn count measures because the variance of the ICCs between the humans and the LENA system was high.

In Chapter 4, the current dissertation presented an exploratory study using the LENA system to investigate environmental effects on child vocalization and conversation counts. The study was conducted in three early intensive behavioral intervention (EIBI) centers in the

Midwest that enrolled 21 young children with ASD. In addition, 32 behavior technicians (BTs) that were employed in the centers were recruited to collect environmental variables such as the location, grouping, delivery method, and objective of on-going interventions in the EIBI centers. The LENA measures and the environmental variables were analyzed using a Linear Mixed Model (LMM) analysis. The LMM results indicated that child vocalizations were significantly higher in the inclusion classroom ($\beta = 0.55$; $p < .001$) compared to the self-contained EIBI center, during group instruction ($\beta = 0.27$; $p < .001$) compared to individual instruction, and when the intervention was delivered naturally ($\beta = 0.26$; $p < .001$) compared to discrete delivery. Although the results showed that environmental variables outside the planned intervention may positively affect child vocalizations, the nature of the vocalizations are unknown because this study did not examine the context of the child's language. As such, additional research is needed to draw a firm interpretation of the implications regarding the environmental effects on the expressive language of children with ASD.

Automated Data Collection for ABA

Interventions based on ABA have shown to be highly effective for children with ASD (Green et al., 2006; Reichow & Wolery, 2009; Wong et al., 2015). Typically, ABA research relies on human observers to collect and measure behavior (Cooper, Heron, & Heward, 2007). Physical limitations that accompany human observation can lead to inaccuracy, observer drift, and bias in measurement (Gast & Ledford, 2014; Kazdin, 2011). Reliable and precise measurement of behavior is especially important for children with ASD because consistent and systematic intervention can increase effectiveness of the implemented intervention (Fisher & Meyer, 2002; Whitaker, 2004). Using automated data collection apparatuses can increase accuracy, validity, and reliability in measurement (Crowley-Koch & Van Houten, 2013); and

consequently, improve and inform current ABA interventions for children with ASD.

Many apparatuses reviewed in Chapter 2 can provide ideas for ABA research involving children with ASD. Researchers investigating joint attention, eye-gaze, or related social skills for children with ASD can use sensors that track eye gazes or facial expressions (e.g., Lancioni et al., 2011; Miller, Wyatt, Casey, & Smith, 2018). The sensors can provide more accurate and reliable data than human observers that may try to detect subtle movements of facial muscles through direct or indirect observations. Also, researchers studying effective prompting sequences or delayed feedback procedures for children with ASD (e.g., Saunders & Saunders, 2012; Yu, Moon, Oah, & Lee, 2013) can utilize microswitches or sensors to collect data. Microswitches or sensors will allow researchers to provide the respective prompt or feedback with increased precision. Whereas human data collectors may struggle to collect behavioral data, measure temporal data for latency or duration, and provide feedback or prompting simultaneously, these automated apparatuses will perform the tasks accurately and without delay.

Furthermore, researchers can use automated data collection apparatuses to measure behaviors in applied settings for larger samples that would have been difficult with human data collection. For example, many researchers recommend collecting repeated, prolonged language samples from the children's natural environment to represent accurate language skill levels in children with ASD (Kasari, Brady, Lord, & Tager-Flusberg, 2013; Sandbank & Yoder, 2014). Although collecting and analyzing natural language data may be possible for studies with few participants, a study that investigates language with large samples of children with ASD may require as many human data collectors as the number of participants. In addition, the collected data for the larger study will need additional human resources for aggregation and analysis of data. An automated data collection apparatus that can collect language data, such as the LENA

system will allow researchers to collect natural language samples while providing automatically aggregated language measures.

The study presented in Chapter 4 is one example where the researcher was able to collect additional data by using the LENA system. Using the LENA system allowed the primary investigator to collect three day-long natural language samples and some details about the respective language environment from 21 children with ASD across three sites. Without the automated data collection system, the primary investigator would have had to use human data collection and considerably reduce the number of participants to match available resources. Although a similar study with a smaller sample size may still present interesting results and implications, the interpretation may have been limited to individual participant effects pertaining to between-child differences. However, by increasing the sample size with the automated data collection, the primary investigator was able to conduct complex statistical analyses that provided preliminary results regarding the effects of environmental variables for global implications.

Reliability of Automated Data Collection Apparatuses

Although precision of measurement is important in ABA, another important dimension of ABA is that the research result is socially valid and meaningful to the participant (Cooper et al., 2007; Wolf, 1978). Thus, using an automated data collection apparatus to increase precision and efficiency in measurement is pointless if the collected data are not valid. Because of this reason, it is crucial that the automated data collection apparatus is reliable for the target population. Researchers should conduct reliability tests for automated data collection apparatuses that are produced as mass-market products and calibration tests for programmable automated data collection apparatuses to make sure that the automated data collection apparatus is reliable to use

for the participants in their study.

Studies examined in Chapter 2 included research that conducted calibration or reliability analysis for the automated data collection apparatus to investigate whether the device was usable for certain populations. Van Camp and Berth (2018) tested the reliability of *Fitbits* for use in children. The *Fitbit* is a mass-consumer product that is calibrated for adults (Van Camp & Berth, 2018). Van Camp and Berth (2018) found that the *Fitbit* may not be reliable to assess physical activity with children who may engage in vertical movement (e.g., climbing) for exercise in playgrounds or parks. In another study, Lancioni and colleagues (2011) conducted research on using microswitch controlled cameras to measure facial expressions in children with severe multiple disabilities. Lancioni and colleagues (2011) pointed out that the microswitch-controlled cameras need multiple calibrations and algorithm tests to ensure that this type of automated data collection apparatus will reliably measure facial expressions of the target participants.

Similarly, Chapter 3 investigated the reliability of the LENA system for language studies involving children with ASD. The LENA system was originally designed for typically developing children (Xu, Yapanel, Gray, 2009). Therefore, there remains a possibility that the LENA system may not be able to reliably measure expressive language of children with ASD who may display atypical language such as stereotypical and echoic prelinguistic vocalizations. The results from Chapter 3 supported the reliability of the LENA system's child vocalization measures for children with ASD. However, conversational turn counts showed high variance with an ICC range of .97 to .04 between the LENA system and human coders. Because the post-hoc analysis did not find a correlation between the child's age or severity of ASD and the ICC, there is still much to learn about why there is high discrepancy in the children's conversational turn counts between the LENA system and human coders. A future reliability study that includes

a qualitative analysis may help reveal child characteristic or environmental factors not identified in Chapter 3 that affect the reliability of the LENA system. This additional information will help researchers make informed decisions on whether to use the LENA system for certain populations.

Study results from both Chapter 2 and 3 suggest that although automated data collection can improve and inform data, the use of the apparatus should be appropriate for both the participants and the research question. For example, if the LENA system was to be used to collect data to assess a focused language intervention (see Chapter 4 for definition) for children with ASD, the frequency measures supplied by the LENA system will not represent whether the child acquired the target language unit. Also, researchers who are interested in the conversations between the children with ASD and adults may use the conversational turn counts supplied by the LENA system but will also need to supplement the LENA data with content analysis conducted by human coders to understand the interaction elements of the conversation. But for a study examining the quantity of vocalizations from a large sample of children, the LENA system may not only help researchers collect language data more accurately and efficiently, but also allow researchers to collect environmental data such as adult language (e.g., Burgess, Audet, & Harjusola-Webb, 2013) within fixed resources.

For Chapter 4, the primary investigator utilized this benefit and collected language measures with the LENA system with respective environment data. Researchers suggest certain environmental variables may positively affect on-going interventions and manipulating those variables may increase the effectiveness of certain interventions (Boyd et al., 2014; Kane et al., 2011). However, collecting environmental variables concurrently with language data throughout the day from a large number of children with ASD requires a lot of time and human resources.

Because the LENA system provided the dependent language measures (i.e., child vocalization and conversational turn counts) automatically, the primary investigator was able to utilize available human resources to collect the independent environmental variables (i.e., location, grouping, delivery method, and objective of intervention) simultaneously. Furthermore, the reliability of the LENA system was established in Chapter 3, so this automated data collection apparatus was appropriate for both the research question and the target population.

Although Chapter 4 provided preliminary information about the relationship between some environmental variables and some forms of expressive language in children with ASD, the results of both Chapters 3 and 4 repeatedly call for additional qualitative research. Chapter 3 found that the reliability of the conversational turn measure was lower than child vocalization counts. And Chapter 4 called for additional qualitative analysis regarding the nature of the increases in child vocalizations under specific environmental variables. Future researchers that seek to use the LENA system for research regarding children with ASD should consider adding human coders for additional qualitative analyses that can inform the quantitative measures collected by the LENA system.

Environmental Effects on Children's Language

The LENA system helped to provide quantitative evidence that children with ASD display significantly higher vocalizations under different environmental variables. Child vocalizations were significantly higher in the inclusion classroom compared to the self-contained EIBI classroom, higher during group instruction compared to individual instruction, and higher during natural interventions compared to discrete interventions. Currently, there are no studies that discuss the quantity of expressive language children with ASD show under different environmental variables. By using an automated data collection device that has shown reliability

for use in the target population, the primary investigator was able to provide preliminary quantitative evidence on the relationship between different environmental variables and children with ASD's expressive language.

Many researchers continue to debate whether inclusion or self-contained, group or individual, or natural or discrete instruction is more effective for language development in children with ASD. Some researchers state that some children with ASD cannot learn through observational learning and/or spontaneous social interactions and need systematic individualized instruction (Fisher & Meyer, 2002; Whitaker, 2004). Others state that since children with ASD only display language in the context of how they acquired the skill, they need to learn language in natural settings and through interaction with typically developing peers (LeBlanc, Dillon, & Sautter, 2009; Schreibman et al., 2015). The exploratory analysis conducted by the LENA system in Chapter 4 may justify the need for a closer investigation of the quality and content of expressive language in children with ASD under different environment variables. Future research in this area could provide additional information that has not yet been considered by previous research on what environment, at what time, is most effective for language development in children with ASD. In addition, identifying environment variables that strengthen existing evidence-based language interventions may be an efficient way to increase the potency of intervention and further facilitate long-term language development for children with ASD.

Results of Chapter 4 also found no significant relationships between conversational turn counts and the environmental variables. Because the study presented in Chapter 3 found that the reliability of the conversational turn measure was lower than child vocalization counts, there is a possibility, however remote, that the conversational turn counts supplied by the LENA system in Chapter 4 were not representative of the children's actual conversation frequencies. If so, further

research is needed to identify an optimal method to measure the number of true interactive conversations in children with ASD. Social communication, with language, is an area that children with ASD commonly show difficulty (American Psychiatric Association, 2013). Identifying possible environmental variables that promote social communication in children with ASD may be another area of future research to benefit language development in children with ASD.

Language Research in Children with ASD

Language is important for children with ASD, as it is a strong predictor for academic success and independent adult life (Roux, Shattuck, Rast, Rava, & Anderson, 2015). Despite continuous research, an estimated 30% of individuals with ASD remain non- or minimally verbal throughout life (Tager-Flusberg & Kasari, 2013). Using an automated data collection apparatus such as the LENA system can provide new information that can inform current language interventions and characteristics research in children with ASD. Researchers should utilize more innovative methods that combine both the qualitative and quantitative dimensions of language to further investigate the implications presented in the current dissertation.

APPENDICES

APPENDIX A

Environmental Data Sheet

The Behavior Technician (BT) data sheet was created by the primary investigator. BTs collected data every time a new activity began for a child with ASD enrolled in Chapter 4. The start time of the activity was entered in the Time column. The BTs circled the appropriate description for the Location, Grouping, Method, and Learning Objective columns. The BTs also recorded a simple description of the activity (e.g., Lunch, inclusion story time) in the Comments column.

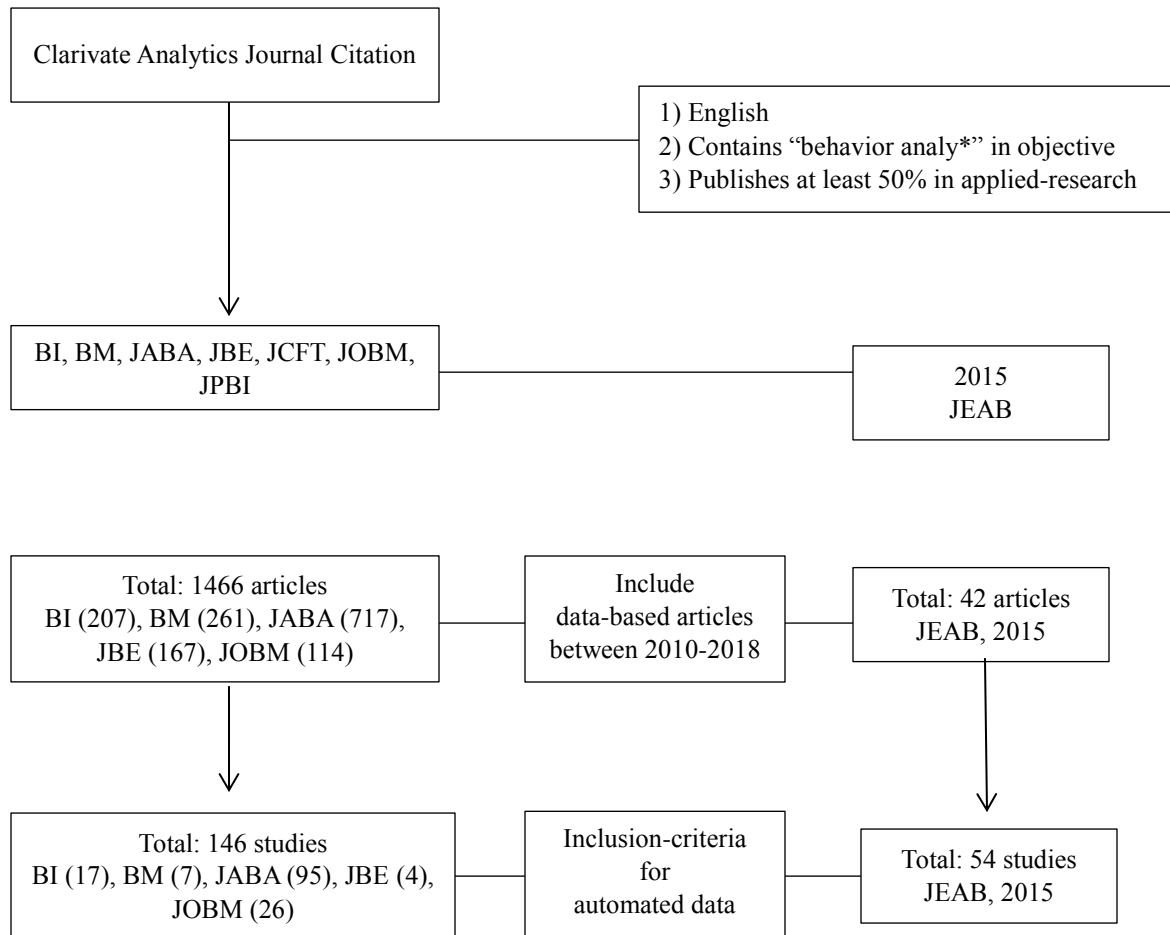
BT Name: _____ Child Number: _____ Date: _____ Page: _____ of _____

[illegible]

Note. E = EIBI, I = inclusion. 1:1 = individual, G = group. DTT = discrete.

APPENDIX B
Flowchart of Study Inclusion and Exclusion

Flowchart of the inclusion and exclusion process of the journals and studies included in this review



REFERENCES

REFERENCES

- Akande, A. (2000). Assessing color identification in children with autism. *Early Child Development and Care, 164*, 95-104. doi:10.1080/0300443001640108
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.) Arlington, VA: American Psychiatric Association.
- Baer, D. M., Montrose, M. W., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 1*, 91-97. doi:10.1901/jaba.1968.1-91
- Bak, M. Y. S., Plavnick, J. B., & Byrne, S. M. (2019). Vocalizations of minimally verbal children with autism spectrum disorder across the school year. *Autism, 23*, 371-382. doi:10.1177/1362361317747576
- Barbarese, W. J., Colligan, R. C., Weaver, A. L., & Katusic, S. K. (2009). The incidence of clinically diagnosed versus research-identified autism in Olmsted county, Minnesota, 1976-1997: Results from a retrospective, population-based study. *Journal of Autism and Developmental Disorders, 39*, 464-470. doi:10.1007/s10803-008-0645-8
- Baril, E. M., & Humphreys, B. P. (2017). An evaluation of the research evidence on the Early State Denver Model. *Journal of Early Intervention, 39*, 321-338. doi:10.1177/1053815117722618
- Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A., & Escobar, G. (2014). Big data in health care: Using analytics to identify and manage high-risk and high-cost patients. *Health Affairs, 33*, 1123-1131. doi:10.1377/hlthaff.2014.0041
- Bekker, M. J., Cumming, T. D., Osborne, N. K., Bruining, A. M., McClean, J. I., & Leland, L. S. (2010). Encouraging electricity savings in a university residential hall through a combination of feedback, visual prompts, and incentives. *Journal of Applied Behavior Analysis, 43*, 327-331. doi:10.1901/jaba.2010.43-327
- Bonato, P. (2005). Advances in wearable technology and applications in physical medicine and rehabilitation. *Journal of NeuroEngineering and Rehabilitation, 2*, 2-5. doi:10.1186/1743-0003-2-2
- Boyd, B. A., Hume, K., McBee, M. T., Alessandri, M., Gutierrez, A., Johnson, L., ... Odom, S. L. (2014). Comparative efficacy of LEAP, TEACCH and non-model-specific special education programs for preschoolers with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 44*, 366-380. doi:10.1007/s10803-013-1877-9

- Burgess, S., Audet, L., & Harjusola-Webb, S. (2013). Quantitative and qualitative characteristics of the school and home language environments of preschool-aged children with ASD. *Journal of Communication Disorders*, 46, 428-439. doi:10.1016/j.jcomdis.2013.09.003
- Caskey, M., Stephens, B., Tucker, R., & Vohr, B. (2014). Adult talk in the NICU with preterm infants and developmental outcomes. *Pediatrics*, 133, 1-7. doi:10.1542/peds.2013-0104
- Chen, C. L. P., & Zhang, C. (2014). Data-intensive applications, challenges, techniques, and technologies: A survey on Big Data. *Information Sciences*, 275, 314-347. doi:10.1016/j.ins.2014.01.015
- Chok, J. T., Demanche, J., Kennedy, A., & Studer, L. (2010). Utilizing physiological measures to facilitate phobia treatment with individuals with autism and intellectual disability: A case study. *Behavioral Interventions*, 25, 325-337. doi:10.1002/bin.312
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6, 284-290. doi:10.1037/1040-3590.6.4.284
- Cnaan, A., Laird, N. M., & Slasor, P. (1997). Tutorial in biostatistics: Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Statistics in Medicine*, 16, 2349-2380. Retrieved from: https://fhs.mcmaster.ca/anesthesiaresearch/documents/1997Avital_CnaanUsingthegenerallinear mixed model to analyse unbalanced.pdf
- Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Upper Saddle River, NJ: Pearson Education.
- Critchfield, T. S., (2014). Online equivalence-based instruction about statistical inference using written explanation instead of match-to-sample training. *Journal of Applied Behavior Analysis*, 47, 606-611. doi:10.1002/jaba.150
- Crowley-Koch, B. J., & Van Houten, R. (2013). Automated measurement in applied behavior analysis: A review. *Behavioral Interventions*, 28, 225-240. doi:10.1002/bin.1366
- Dallery, J., Raiff, B. R., & Grabinski, M. J. (2013). Internet-based contingency management to promote smoking cessation: A randomized controlled study. *Journal of Applied Behavior Analysis*, 46, 750-764. doi:10.1002/jaba.89
- Darling, K. E., Fahrenkamp, A. J., Wilson, S. M., Karazsia, B. T., & Sato, A. F. (2017). Does social support buffer the association between stress eating and weight gain during the transition to college? Differences by gender. *Behavior Modification*, 41, 368-381. doi:10.1177/0145445516683924

- Dawson, G., Rogers, S., Munson, J., Smith, M., Winter, J., Greenson, J., ... Verley, J. (2010). Randomized controlled trial of an intervention for toddlers with autism: The Early Start Denver Model. *Pediatrics*, 125, e17-e23. doi:10.1542/peds.2009-0958
- Dettmer, S., Simpson, R. L., Myles, B. S., & Ganz, J. B. (2000). The use of visual supports to facilitate transitions of students with autism. *Focus on Autism and other Developmental Disabilities*, 15, 163-169. doi:10.1177/108835760001500307
- Dotson, W. H., Leaf, J. B., Sheldon, J. B., & Sherman, J. A. (2010). Group teaching of conversational skills to adolescents on the autism spectrum. *Research in Autism Spectrum Disorders*, 4, 199-209. doi:10.1016/j.rasd.2009.09.005
- Dykstra, J. R., Sabatos-DeVito, M. G., Irvin, D. W., Boyd, B. A., Hume, K. A., & Odom, S. L. (2012). Using the Language Environment Analysis (LENA) system in preschool classrooms with children with autism spectrum disorders. *Autism*, 17, 582-594. doi:10.1177/1362361312446206
- Eigsti, I., de Marchena, A. B., Schuh, J. M., & Kelley, E. (2011). Language acquisition in autism spectrum disorders: A developmental review. *Research in Autism Spectrum Disorders*, 5, 681-691. doi:10.1016/j.rasd.2010.09.001
- Estes, A., Munson, J., Rogers, S. J., Greenson, J., Winter, J., & Dawson, G. (2015). Long-term outcomes of early intervention in 6-year-old children with autism spectrum disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, 54, 580-587. doi:10.1016/j.jaac.2015.04.005
- Fahmie, T. A., Macaskill, A. C., Kazemi, E., Elmer, U. C. (2018). Prevention of the development of problem behavior: A laboratory model. *Journal of Applied Behavior Analysis*, 51, 25-39. doi:10.1002/jaba.426
- Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., ... Reilly, J. S. (1993). *MacArthur communicative development inventory: Users guide and technical manual*. San Diego, CA: Singular Publishing Company.
- Fienup, D. M., Covey, D. P., & Critchfield, T. S. (2013). Teaching brain-behavior relations economically with stimulus equivalence technology. *Journal of Applied Behavior Analysis*, 43, 19-33. doi:10.1901/jaba.2010.43-19
- Finn, M., Barnes-Holmes, D., Hussey, I., & Graddy, J. (2016). Exploring the behavioral dynamics of the implicit relational assessment procedure: The impact of three types of introductory rules. *Psychological Record*, 66, 309-321. doi:10.1007/s40732-016-0173-4
- Fisher, M., & Meyer, L. H. (2002). Development and social competence after two years for students enrolled in inclusive and self-contained educational programs. *Research and*

- Practice for Persons with Severe Disabilities*, 27, 165-174. doi:10.2511/rpsd.27.3.165
- Fletcher-Watson, S., & McConachie, H. (2017). The search for an early intervention outcome measurement tool in autism. *Focus on Autism and Other Developmental Disabilities*, 32, 71-80. doi: 10.1177/1088357615583468
- Ford, M., Baer, C. T., Xu, D., Yapanel, U., & Gray, S. (2008). *The LENA language environment analysis system: Audio specifications of the DLP-0121*. Boulder, CO: LENA Foundation. Retrieved from https://3ezaxq2cvfwhsrafg2qaq2p4-wpengine.netdna-ssl.com/wp-content/uploads/2016/07/LTR-03-2_Audio_Specifications.pdf
- Galbraith L. A., & Normand, M. P. (2017). Step it up! Using the good behavior game to increase physical activity with elementary school students at recess. *Journal of Applied Behavior Analysis*, 50, 856-860. doi:10.1002/jaba.402
- Gast, D. L., & Ledford, J. R. (2014). *Single case research methodology: Applications in special education and behavioral sciences* (2nd ed.). New York, NY: Routledge.
- Gernsbacher, M. A., Morson, E. M., & Grace, E. J. (2016). Language and speech in autism. *Annual Review of Linguistics*, 2, 413-425. doi:10.1146/annurev-linguistics-030514-124824
- Gilkerson, J., Coulter, K. K., & Richards, J. A. (2008). *Transcriptional analyses of the LENA natural language corpus*. Boulder, CO: LENA Foundation. Retrieved from https://3ezaxq2cvfwhsrafg2qaq2p4-wpengine.netdna-ssl.com/wp-content/uploads/2016/07/LTR-06-2_Transcription.pdf
- Gilkerson, J., & Richards, J. A. (2009). *The power of talk* (2nd ed.). Boulder, CO: LENA Foundation. Retrieved from https://www.lena.org/wp-content/uploads/2016/07/LTR-01-2_PowerOfTalk.pdf
- Golfeto R. M., & de Souza, D. G. (2015). Sentence production after listener and echoic training by prelingual deaf children with cochlear implants. *Journal of Applied Behavior Analysis*, 48, 363-378. doi:10.1002/jaba.197
- Grindle, C. F., Hughes, J. C., Saville, M., Huxley, K., & Hastings, R. P. (2013). Teaching early reading skills to children with autism using MimioSprout Early Reading. *Behavioral Interventions*, 28, 203-224. doi:10.1002/bin.1364
- Green, V. A., Pituch, K. A., Itchon, J., Choi, A., O'Reilly, M., & Sigafoos, J. (2006). Internet survey of parents of children with autism. *Research in Developmental Disabilities*, 27, 70-84. doi:10.1016/j.ridd.2004.12.002
- Greer, R. D., & Ross, D. E. (2008). *Verbal behavior analysis: Inducing and expanding new verbal capabilities in children with language delays*. Boston, MA: Allyn and Bacon.

- Hankla, M. E., Kohn, C. S., & Normand, M. P. (2018). Teaching college students to pour accurately using behavior skills training: Evaluation of the effect of peer modeling. *Behavioral Interventions*, 33, 136-149. doi:10.1002/bin.1509
- Heck, R. H. (2001). Multilevel modeling with SEM. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 89-127). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Hirst, J. M., DiGennaro Reed, F. D., & Reed, D. D. (2013). Effects of varying feedback accuracy on task acquisition: A computerized translational study. *Journal of Behavioral Education*, 22, 1-15. doi:10.1007/s10864-012-9162-0
- Jamison, C. A., Kelley, D. P. III, Schmitt, C., Harvey, M. T., Harvey, A. C., & Meyer, E. (2014). Impact of an overt response system on staff performance and maintenance of computer-based instruction. *Journal of Organizational Behavior Management*, 34, 279-290. doi:10.1080/01608061.2014.973630
- Ingersoll, B., & Schreibman, L. (2006). Teaching reciprocal imitation skills to young children with autism using a naturalistic behavioral approach: Effects on language, pretend play, and joint attention. *Journal of Autism and Developmental Disorders*, 36, 487-505. doi:10.1007/s10803-006-0089-y
- Interagency Autism Coordinating Committee. (2014). IACC Strategic plan for autism spectrum disorder research – 2013 update. *U.S. Department of Health and Human Services Interagency Autism Coordinating Committee*. Retrieved from <http://iacc.hhs.gov/strategicplan/2013/index.shtml>
- Jang, J., Dixon, D. R., Tarbox, J., & Granpeesheh, D. (2011). Symptom severity and challenging behavior in children with ASD. *Research in Autism Spectrum Disorders*, 5, 1028-1032. doi:10.1016/j.rasd.2010.11.008
- Kaiser, A. P., Hancock, T. B., & Nietfeld, J. P. (2000). The effects of parent-implemented enhanced milieu teaching on the social communication of children with have autism. *Early Education and Development*, 11, 423-446. doi:10.1207/s15566935eed1104_4
- Kane, M., Connell, J. E., & Pellecchia, M. (2010) A quantitative analysis of language interventions for children with autism. *Behavior Analyst Today*, 11, 128-144. doi:10.1037/h0100696
- Kasari, C., Brady, N., Lord, C., & Tager-Flusberg, H. (2013). Assessing the minimally verbal school-aged children with autism spectrum disorder. *Autism Research*, 6, 479-493. doi:10.1002/aur.1334
- Kazbour, R. R., & Bailey, J. S. (2010). An analysis of a contingency program on designated

- drivers at a college bar. *Journal of Applied Behavior Analysis*, 43, 273-277. doi:10.1901/jaba.2010.43-273
- Kazdin, A. E. (1979). Unobtrusive measures in behavioral assessment. *Journal of Applied Behavior Analysis*, 12, 713-724. doi:10.1901/jaba.1979.12-713
- Kazdin, A. E. (2011). *Single-case research design: Methods for clinical and applied settings* (2nd ed.). New York, NY: Oxford University Press, Inc.
- Kelley, M. E., Liddon, C. J., Ribeiro, A., Grief, A. E., & Podlesnik, C. A. (2015). Basic and translational evaluation of renewal of operant responding. *Journal of Applied Behavior Analysis*, 48, 390-401. doi:10.1002/jaba.209
- Kelly, M. B. (1977). A review of the observational data-collection and reliability procedures reported in the Journal of Applied Behavior Analysis. *Journal of Applied Behavior Analysis*, 10, 97-101. doi:10.1901/jaba.1977.10-97
- Koegel, L. K., Koegel, R. L., Harrower, J. K., & Carter, C. M. (1999). Pivotal response intervention I: Overview of approach. *Research and practice for persons with severe disabilities*, 24, 174-185. doi:10.2511/rpsd.24.3.174
- Kogen, M. D., Vladutiu, C. J., Schieve, L. A., Ghandour, R. M., Blumberg, S. J., Zablotsky, B., ... Lu, M. C. (2018). The prevalence of parent-reported autism spectrum disorder among US children. *Pediatrics*, 142(6), 1-11. doi:10.1542/peds.2017-4161
- Kwok, E. Y. L., Brown, H. M., Smyth, R. E., & Cardy, J. O. (2014). Meta-analysis of receptive and expressive language skills in autism spectrum disorder. *Research in Autism Spectrum Disorders*, 9, 202-222. doi:10.1016/j.rasd.2014.10.008
- Lancioni, G. E., Bellini, D., Oliva, D., Singh, N. N., O'Reilly, M. F., Lang, R., & Didden, R. (2011). Camera-based microswitch technology to monitor mouth, eyebrow, and eyelid responses of children with profound multiple disabilities. *Journal of Behavioral Education*, 20, 4-14. doi:10.1007/s 10864-010-9117-2
- Lancioni, G. E., Singh, N. N., O'Reilly, M. F., Sigafoos, J., D'Amico, F., Addante, L. M., & Pinto, K. (2017). Persons with advanced Alzheimer's disease engage in mild lag exercise supported by technology-aided stimulation and prompts. *Behavior Modification*, 41, 3-20. doi:10.1177/0145445516649581
- Larson, T. A., Normand, M. P., & Hustyi, K. M. (2011). Preliminary evaluation of an observation system for recording physical activity in children. *Behavioral Interventions*, 26, 193-203. doi:10.1002/bin.332
- Le Couteur, A., Lord, C., & Rutter, M. (2003). *Autism diagnostic interview-revised (ADI-R)*. Los Angeles, CA: Autism Genetic Resource Exchange.

- LeBlanc, L.A., Dillon, M., & Sautter, R. A. (2009) Establishing mand and tact repertoires. In R. A. Rehfeldt & Y. Barnes-Holmes (Eds.) *Derived relational responding: Applications for learners with autism and other developmental disabilities* (pp. 79-108). Oakland, CA: New Harbinger.
- LENA Research Foundation. (2015). <https://www.lena.org>
- Lord, C., Rutter, M., DiLavorne, P. C., Risi, S., Gotham, K., & Bishop, S. L. (2012). *Autism diagnostic observation schedule, second edition (ADOS-2)*. Torrance, CA: Western Psychological Services.
- Lowe, S. A., & ÓLaighin, G. (2014). Monitoring human health behavior in one's living environment: A technological review. *Medical Engineering & Physics*, 36, 147-168. doi:10.1016/j.medengphy.2013.11.010
- Luyster, R. L., Kadlec, M. B., Carter, A., & Tager-Flusberg, H. (2008). Language assessment and development in toddlers with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 38, 1426-1438. doi:10.1007/s10803-007-0510-1
- Lyons, E. J., Lewis, Z. H., Mayrsohn, B. G., & Rowland, J. L. (2014). Behavior change techniques implemented in electronic lifestyle activity monitors: A systematic content analysis. *Journal of Medical Internet Research*, 16, e192. doi:10.2196/jmir.3469
- MacDonald, R., Parry-Cruwys, D., Dupere, S., & Ahearn, W. (2014). Assessing progress and outcome of early intensive behavioral intervention for toddlers with autism. *Research in Developmental Disabilities*, 35, 3632-3644. doi:10.1016/j.ridd.2014.08.036
- Mahon, C., Lyddy, F., & Barnes-Holmes, D. (2010). Recombinative generalization of subword units using matching to sample. *Journal of Applied Behavior Analysis*, 43, 303-307. doi:10.1901/jaba.2010.43-303
- Makrygianni, M. K., & Reed, P. (2010). A meta-analytic review of the effectiveness of behavioural early intervention programs for children with autistic spectrum disorders. *Research in Autism Spectrum Disorders*, 4, 577-593. doi:10.1016/j.rasd.2010.01.014
- Matson, J. L., & Boisjoli, J. A. (2009). The token economy for children with intellectual disability and/or autism: A review. *Research in Developmental Disabilities*, 30, 240-248. doi:10.1016/j.ridd.2008.04.001
- Matson, J. L., & LoVullo, S. V. (2008). A review of behavioral treatments for self-injurious behavior of persons with autism spectrum disorders. *Behavior Modification*, 32, 61-76. doi:10.1177/0145445507304581

- McDonell, M. G., Howell, D. N., McPherson, S., Cameron, J. M., Srebnik, D., Roll, J. M., & Ries, R. K. (2012). Voucher-based reinforcement for alcohol abstinence using the ethyl-glucuronide alcohol biomarker. *Journal of Applied Behavior Analysis*, 45, 161-165. doi:10.1901/jaba.2012.45-161
- McLeish, A. C., Luberto, C. M., & O' Bryan, E. M. (2016). Anxiety sensitivity and reactivity to asthma-like sensations among young adults with asthma. *Behavior Modification*, 40, 164-177. doi:10.1177/0145445515607047
- Mullen, E. M. (1995). *Mullen scales of early learning* (AGS ed.). Los Angeles, CA: Western Psychological.
- Napolitano, M. A., Lloyd-Richardson, E. E., Fava, J. L., & Marcus, B. H. (2011). Targeting body image schema for smoking cessation among college females: Rationale, program description, and pilot study results. *Behavior Modification*, 35, 323-346. doi:10.1177/0145445511404840
- National Autism Center. (2015). *Findings and conclusions: National standards project, phase 2*. Randolph, MA: Author. Retrieved from: <http://www.nationalautismcenter.org/090605-2/>
- National Research Council. (2001). *Educating children with autism*. Washington, DC: The National Academies Press. doi:10.17226/10017.
- Norrelgen, F., Fernell, E., Eriksson, M., Hedvall, Å., Persson, C., Sjölin, M., ... Kjellmer, L. (2015). Children with autism spectrum disorders who do not develop phrase speech in the preschool years. *Autism*, 19, 934-943. doi:10.1177/1362361314556782
- Odom, S. L., Boyd, B. A., Hall, L. J., & Hume, K. (2010). Evaluation of comprehensive treatment models for individuals with autism spectrum disorders. *Journal of Autism and Developmental Disabilities*, 40, 425-436. doi:10.1007/s10803-009-0825-1
- Olejnik, S., & Algina, J. (2000). Measures of effect size for comparative studies: Applications, interpretations, and limitations. *Contemporary Educational Psychology*, 25, 241-286. doi:10.1006/ceps.2000.1040
- Oliveira-Castro, J. M., Foxall, G. R., & Wells, V. K. (2010). Consumer brand choice: Money allocation as a function of brand reinforcing attributes. *Journal of Organizational Behavior Management*, 30, 161-175. doi:10.1080/01608061003756455
- Penrod, B., Wallace, M. D., Reagon, K., Betz, A., & Higbee, T. S. (2010). A component analysis of a parent-conducted multi-component treatment for food selectivity. *Behavioral Interventions*, 25, 207-228. doi:10.1002/bin.307

- Peters-Scheffer, N., Didden, R., Korzilius, H., & Sturmey, P. (2011). A meta-analytic study on the effectiveness of comprehensive ABA-based early intervention programs for children with autism spectrum disorder. *Research in Autism Spectrum Disorders*, 5, 60-69. doi:10.1016/j.rasd.2010.03.011
- Peterson, K. M., Piazza, C. C., & Volkert, V. M. (2016). A comparison of a modified sequential oral sensory approach to an applied behavior-analytic approach in the treatment of food selectivity in children with autism spectrum disorder. *Journal of Applied Behavior Analysis*, 49, 485-511. doi:10.1002/jaba.332
- Pickles, A., Anderson, D. K., & Lord, C. (2014). Heterogeneity and plasticity in the development of language: A 17-year follow-up of children referred early for possible autism. *Journal of Child Psychology and Psychiatry*, 55, 1354-1362. doi:10.1111/jcpp.12269
- Plavnick, J. B., Thompson, J. L., Englert, C. S., Mariage, T., & Johnson, K. (2016). Mediating access to Headsprout® Early Reading for children with autism spectrum disorders. *Journal of Behavioral Education*, 25, 357-378. doi:10.1007/s10864-015-9244-x
- Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big Data*, 1, 51-59. doi:10.1089/big.2013.1508
- Rankine, J., Li, E., Lurie, S., Rieger, H., Fourie, E., Siper, P. M., ... Kolevzon, A. (2017). Language ENvironment Analysis (LENA) in Phelan-McDermid syndrome: Validity and suggestions for use in minimally verbal children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 47, 1605-1617. doi: 10.1007/s10803-017-3082-8
- Raudenbush, S. W. (1988). Methodological advances in analyzing the effects of schools and classrooms on student learning. *Review of Research in Education*, 15, 423-475. doi:10.3102/0091732X015001423
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). London, England: Sage.
- Reichle, J., Dropik, P. L., Alden-Anderson, E., & Haley, T. (2008). Teaching a young child with autism to request assistance conditionally: A preliminary study. *American Journal of Speech-Language Pathology*, 17, 231-240. doi:10.1044/1058-0360
- Reichow, B., & Sabornie, E. J. (2009). Brief report: Increasing verbal greeting initiations for a student with autism via a Social Story™ intervention. *Journal of Autism and Developmental Disorders*, 39, 1740-1743. doi:10.1007/s10803-009-0814-4
- Reichow, B., & Wolery, M. (2009). Comprehensive synthesis of early intensive behavioral interventions for young children with autism based on the UCLA Young Autism Project

- model. *Journal of Autism and Developmental Disabilities*, 39, 23-41.
doi:10.1007/s10803-008-0596-0
- Repp, A. C., & Felce, D. (1990). A microcomputer system used for evaluative and experimental behavioural research in mental handicap. *Journal of Applied Research in Intellectual Disabilities*, 3, 21-32. doi:10.1111/j.1468-3148.1990.tb00078.x
- Reyes, J. R., Vollmer, T. R., & Hall, A. (2011). Replications and extensions in arousal assessment for sex offenders with developmental disabilities. *Journal of Applied Behavior Analysis*, 44, 369-373. doi:10.1901/jaba.2011.44-369
- Ribeiro, D. M., Miguel, C. F., & Goyos, C. (2015). The effects of lister training on discriminative control by elements of compound stimuli in children with disabilities. *Journal of the Experimental Analysis of Behavior*, 104, 48-62. doi:10.1002/jeab.161
- Roose K. M., & Williams, W. L. (2018). An evaluation of the effects of very difficult goals. *Journal of Organizational Behavior Management*, 38, 18-48.
doi:10.1080/01608061.2017.1325820
- Roux, A. M., Shattuck, P. T., Rast, J. E., Rava, J. A., & Anderson, K. (2015). *National Autism Indicators Report: Transition into Young Adulthood*. Philadelphia, PA: Life Course Outcomes Research Program, A.J. Drexel Autism Institute, Drexel University. Retrieved from: <http://drexel.edu/autismoutcomes/publications-and-reports/publications/National-Autism-Indicators-Report-Transition-to-Adulthood/#sthash.1f9xWUan.dpbs>
- Romanczyk, R. G., Kent, R. N., Diament, C., & O'Leary, K. D. (1973). Measuring the reliability of observational data: A reactive process. *Journal of Applied Behavior Analysis*, 6, 175-184. doi:10.1901/jaba.1973.6-175
- Saini, V., Fisher, W. W. & Pisman, M. D. (2017). Persistence during and resurgence following noncontingent reinforcement implemented with and without extinction. *Journal of Applied Behavior Analysis*, 50, 377-392. doi:10.1002/jaba.380
- Sandbank, M., Woynaroski, T., Watson, L. R., Gardner, E., Kaysili, B. K., & Yoder, P. (2017). Predicting intentional communication in preverbal preschoolers with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 47, 1581-1594.
doi:10.1007/s10803-017-3052-1
- Sandbank, M., & Yoder, P. (2014). Measuring representative communication in young children with developmental delay. *Topics in Early Childhood Special Education*, 34, 133-141.
doi:10.1177/0271121414528052
- Sanders, E. J., Irvin, D. W., Belardi, K., McCune, L., Boyd, B. A., & Odom, S. L. (2016). The questions verbal children with autism spectrum disorder encounter in the inclusive preschool classroom. *Autism*, 20, 96-105. doi:10.1177/1362361315569744
- Saunders, M. D., & Saunders, R. R. (2012). Teaching individuals to signal for assistance in a

- timely manner. *Behavioral Interventions*, 27, 193-206. doi:10.1002/bin.1346
- Schnell, L. K., Sidener, T. M., Debar, R. M., Vladescu, J. C., & Kang, S. (2018). Effects of computer-based training on procedural modifications to standard functional analyses. *Journal of Applied Behavior Analysis*, 51, 87-98. doi:10.1002/jaba.423
- Schreibman, L., Dawson, G., Stahmer, A. C., Landa, R., Rogers, S. J., McGee, G. G., ... Halladay, A. (2015). Naturalistic developmental behavioral interventions: Empirically validated treatments for autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 45, 2411–2428. doi:10.1007/s10803-015-2407-8
- Schultz, N. R., Kohn, C. S., & Musto, A. (2017). Examination of a multi-element intervention on college students' electricity consumption in on-campus housing. *Behavioral Interventions*, 32, 79-90. doi:10.1002/bin.1463
- Shull, P. B., Jirattigalachote, W., Hunt, M. A., Cutkosky, M. R., & Delp, S. L. (2014). Quantified self and human movement: A review on the clinical impact of wearable sensing and feedback for gait analysis and intervention. *Gait and Posture*, 40, 11-19. doi:10.1016/j.gaitpost.2014.03.189
- Sigurdsson, V., Larsen, N. M., Gunnarsson, D. (2014). Healthy food products at the point of purchase: An in-store experimental analysis. *Journal of Applied Behavior Analysis*, 47, 151-154. doi:10.1002/jaba.91
- Simpson, R. L., De Boer-Ott, S. R., & Smith-Myles, B. (2003). Inclusion of learners with autism spectrum disorders in general education settings. *Topics in Language Disorders*, 23, 116-133
- Simpson, R. L., Mundschenk, N. A., & Heflin, J. (2011). Issues, policies, and recommendations for improving the education of learners with autism spectrum disorders. *Journal of Disability Policy Studies*, 22, 3-17. doi:10.1177/1044207310394850
- Skinner, B. F. (1956). A case history in scientific method. *American Psychologist*, 11, 221-233. doi:10.1037/h0047662
- Sparrow, S., Balla, D., & Cicchetti, D. (1984). *Vineland adaptive behavior scales*. Circle Pines, MN: American Guidance Service.
- Springer, B., Brown, T., & Duncan, P. K. (1981). Current measurement in applied behavior analysis. *Behavior Analyst*, 4, 19-31. doi:10.1007/BF03391849
- Stahmer, A. C., Collings, N. M., & Palinkas, L. A. (2005). Early interventions practices for children with autism: Descriptions from community providers. *Focus on Autism and Other Developmental Disabilities*, 20, 66-79. doi:10.1177/10883576050200020301
- Steingrimsdottir, H. S., & Arntzen, E. (2011). Using conditional discrimination procedures to

- study remembering in an Alzheimer's patient. *Behavioral Interventions*, 26, 179-192. doi:10.1002/bin.334
- Sterponi, L., de Kirby, K., & Shankey, J. (2015). Rethinking language in autism. *Autism*, 19, 517-526. doi:10.1177/1362361314537125
- Storey, C., McDowell, C., & Leslie, J. C. (2017). Evaluating the efficacy of the Headsprout® reading program with children who have spent time in care. *Behavioral Interventions*, 32, 285-293. doi:10.1002/bin.1476
- Suskind, D. L., Leffel, K. R., Graf, E., Hernandez, M. W., Gunderson, E. A., Sapolich, S. G., ... Levine, S. C. (2016). A parent-directed language intervention for children of low socioeconomic status: A randomized controlled pilot study. *Journal of Child Language*, 43, 366-406. doi:10.1017/S0305000915000033
- Swan, M. (2013). The quantified self: Fundamental disruption in big data and biological discovery. *Big Data*, 1, 85-99. doi:10.1089/big.2012.0002
- Tager-Flusberg, H., & Kasari, C. (2013). Minimally verbal school-aged children with autism spectrum disorder: The neglected end of the spectrum. *Autism Research*, 6, 468-478. doi:10.1002/aur.1329
- Tager-Flusberg, H., Rogers, S., Cooper, J., Landa, R., Lord, C., Paul, R., ... Yoder, P. (2009). Defining spoken language benchmarks and selecting measures of expressive language development for young children with autism spectrum disorders. *Journal of Speech, Language, and Hearing Research*, 52, 643-652. doi:10.1044/1092-4388
- Tanji T., & Noro, F. (2011). Matrix training for generative spelling in children with autism spectrum disorder. *Behavioral Interventions*, 26, 326-339. doi:10.1002/bin.340
- Trembath, D., Westerveld, M. F., Teppala, S., Thirumanickam, A., Sulek, R., Rose, V., ... Vivanti, G. (2019). Profiles of vocalization change in children with autism receiving early intervention. *Autism Research*. Advance online publication. doi:10.1002/aur.2075
- Van Camp, C. M., & Berth, D. (2018). Further evaluation of observational and mechanical measures of physical activity. *Behavioral Interventions*, 33, 284-296. doi:10.1002/bin.1518
- van Widenfelt, B. M., Treffers, P. D. A., de Beurs, E., Siebelink, B. M., & Koudijs, E. (2005). Translation and cross-cultural adaptation of assessment instruments used in psychological research with children and families. *Clinical Child and Family Psychology Review*, 8, 135-147. doi:10.1007/s10567-005-4752-1
- VanDam, M., Oller, D. K., Ambrose, S. E., Gray, S., Richards, J. A., Gilkerson, J., ... Moeller, M. P. (2015). Automated vocal analysis of children with hearing loss and their typical and atypical peers. *Ear and Hearing*, 36, 146-152. doi:10.1097/AUD.0000000000000138

- Vanwagner, M., Van Houten, R., & Betts, B. (2011). The effects of a rectangular rapid-flashing beacon on vehicle speed. *Journal of Applied Behavior Analysis*, 44, 629-633. doi:10.1901/jaba.2011.44-629
- Virués-Ortega, J. (2010). Applied behavior analytic intervention for autism in early childhood: Meta-analysis, meta-regression and dose-response meta-analysis of multiple outcomes. *Clinical Psychology Review*, 30, 387-399. doi:10.1016/j.cpr.2010.01.008
- Vohr, B. R., Topol, D., Watson, V., St Pierre, L., & Tucker, R. The importance of language in the home for school-age children with permanent hearing loss. *Acta Paediatrica*, 103, 62-69. doi:10.1111/apa.12441
- Wang, Z., Miller, K., & Cortina, K. S. (2013). Using the LENA in teacher training: Promoting student involvement through automated feedback. *Unterrichtswissenschaft*, 41, 290-305. Retrieved from https://www.researchgate.net/profile/Kai_Cortina/publication/261656154_Using_the_LENA_in_teacher_training_Promoting_student_involvement_through_automated_feedback/links/53ff4afc0cf2da31542dd2d4/Using-the-LENA-in-teacher-training-Promoting-student-involvement-through-automated-feedback.pdf
- Warren, S. F., Gilkerson, J., Richards, J. A., Oller, D. K., Xu, D., Yapanel, U., & Gray, S. (2010). What automated vocal analysis reveals about the vocal production and language learning environment of young children with autism. *Journal of Autism and Developmental Disorders*, 40, 555-569. doi:10.1007/s10803-009-0902-5
- Washington, W. D., Banna, K. M., & Gibson, A. L. (2016). Preliminary efficacy of prize-based contingency management to increase activity levels in healthy adults [Special Issue]. *Journal of Applied Behavior Analysis, Health Psychology and Applied Behavior Analysis*, 231-245. doi:10.1002/jaba.119
- Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, 24, 2143-2152. doi:10.1177/0956797613488145
- Weismer, S. E., Lord, C., & Esler, A. (2010). Early language patterns of toddlers on the autism spectrum compared to toddlers with developmental delay. *Journal of Autism and Developmental Disorders*, 40, 1259-1273. doi:10.1007/s10803-010-1983-1
- Whitaker, P. (2004). Fostering communication and shared play between mainstream peers and children with autism: Approaches, outcomes, and experiences. *British Journal of Special Education*, 31, 215-222. doi:10.1111/j.0952-3383.2004.00357.x
- Wolf, M. M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis*, 11, 203-214.

doi:10.1901/jaba.1978.11-203

- Wong, C., Odom, S. L., Hume, K. A., Cox, A. W., Fettig, A., Kucharczyk, S., ... Schultz, T. R. (2015). Evidence-based practices for children, youth, and young adults with autism spectrum disorder: A comprehensive review. *Journal of Autism and Developmental Disabilities*, 45, 1951-1966. doi:10.1007/s10803-014-2351-z
- Xu, D., Yapanel, U., & Gray, S. (2009). *Reliability of the LENA language environment analysis system in young children's natural home environment*. Boulder, CO: LENA Foundation. Retrieved from https://3ezaxq2cvfwhsrafg2qaq2p4-wpengine.netdna-ssl.com/wp-content/uploads/2016/07/LTR-05-2_Reliability.pdf
- Yoder, P. J., Oller, D. K., Richards, J. A., Gray, S., & Gilkerson, J. (2013). Stability and validity of an automated measure of vocal development from day-long samples in children with and without autism spectrum disorder. *Autism Research*, 6, 103-107. doi:10.1002/aur.1271
- Yu, E., Moon, K., Oah, S., & Lee, Y. (2013). An evaluation of the effectiveness of an automated observation and feedback system on safe sitting postures. *Journal of Organizational Behavior Management*, 33, 104-127. doi:10.1080/01608061.2013.785873
- Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2002). *Preschool language scales* (4th ed.). San Antonio, TX: Psychological Corporation.