

Live coding the Communication Complexity Scale:

Validation of a clinically relevant observation tool

© 2020

By Jordana Mancini
B.A. Emory University, 2018

Submitted to the graduate degree program in Speech-Language Hearing and the Graduate
Faculty of the University of Kansas in partial fulfillment of the requirements
for the degree of Master of Arts.

Chair: Kris Pedersen, SLPD, CCC-SLP

Co-Chair: Nancy C. Brady, PhD

Debora B. Daniels, PhD

Date Defended: July 14, 2020

The thesis committee for Jordana Mancini certifies that this is the approved version of the following thesis:

Live coding the Communication Complexity Scale:
Validation of a clinically relevant observation tool

Chair: Kris Pedersen, SLPD, CCC-SLP

Co-Chair: Nancy C. Brady, PhD

Date Approved: July 20, 2020

Abstract

In this preliminary research study, the effect of live scoring versus video scoring of the observational rating tool, the Communication Complexity Scale (CCS), was examined through three participant variables and one assessment variable: communication level of the participant, age of the participant, severity of autism of the participant, and assessment task function (joint attention or behavior regulation). Participants included 42 children between three and 18 years of age, who were referred by special educators in the suburban Kansas City, KS and Lawrence, KS school districts. Children were administered a 30-minute, scripted play-based assessment which used “sabotaged” games or toys to promote communication of the research subjects. The CCS was used to score the assessment by a live coder, as well as a video coder at a delayed time. Through analysis conducted in this study, it was found that there was not a significant difference when applying the CCS as an observational tool regardless of the participants communication level, age, severity of autism, nor the function of the task assessed by the scripted play-assessment. In clinical application, it was found that live and video scoring are both methodologically sound mediums to apply the CCS, which can be used based on clinician or client preferences or needs.

Acknowledgements

I would like to thank the following people in their support in conducting research and writing of this thesis:

The participants in my research project and their caregivers, for allowing myself and others into their school day to play and learn together, and who without their participation this research would not be possible.

My advisor and chairperson, Nancy Brady, for her continual support in this project and for allowing me to research further aspects of the Measuring Early Communication in Students with Autism (MECSA) project.

Kandace Fleming, for her consistent guidance throughout the thesis process and without whom I could not have conducted such well-thought-out analysis.

All past and present members of the MECSA team, with special consideration to Kris Matthews, for her support as a supervisor and mentor throughout the training and work on the project, and Lisa Hallberg, for her help in organizing the data for such a large research project.

My thesis committee, including Debby Daniels and Kris Pedersen, for their consistent support and willingness to help me succeed throughout my master's program.

The educators of Shawnee Mission, Blue Valley, Olathe, and Lawrence, Kansas, for allowing us to work with students in their schools and conduct research over the past several years.

Finally, to my parents for their constant support and dedication to my education.

Table of Contents

Abstract.....	iii
Acknowledgement.....	iv
List of Figures.....	vii
List of Tables.....	viii
Chapter 1: Review of the Literature.....	1
Development of the Communication Complexity Scale.....	1
Live scoring of assessments.....	5
Participant and assessment variables.....	8
Communication level.....	8
Age.....	9
Severity of autism.....	10
Communication function.....	10
Purpose of the current project.....	11
Research questions.....	12
Chapter 2: Methodology.....	14
Participants.....	14
Demographics.....	14
Measures.....	15
Vineland Adaptive Behavior Scales.....	15
Childhood Autism Rating Scale.....	16
Application of the CCS.....	17
Coder criteria.....	18

General analysis approach.....	19
Chapter 3: Results.....	21
Preliminary descriptive information.....	21
Research Question 1: Communication level of participant.....	22
Research Question 2: Age of participant.....	25
Research Question 3: Severity of autism of participant.....	26
Research Question 4: Function of scripted assessment task.....	26
Chapter 4: Discussion.....	29
Communication level of participant.....	29
Age of participant.....	30
Severity of autism of participant.....	31
Function of scripted assessment task.....	31
Clinical implications.....	32
Limitations.....	33
Future directions.....	34
References.....	35
Appendix.....	40

List of Figures

Figure 1: Race and ethnic demographics of participants.....	15
Figure 2: Distribution of optimal and typical scores in video and live coders.....	21

List of Tables

Table 1: Difference in live versus video optimal and typical scores.....	22
Table 2: Optimal score communication level descriptive statistics.....	24
Table 3: Typical score communication level descriptive statistics.....	24
Table 4: Comparison of the difference of task score means between live and video coders.....	27
Table 5: Task score means and effect sizes by task.....	28

Chapter 1: Review of the Literature

Development of the Communication Complexity Scale

Within the last decade, the gap in readily available, reliable assessment tools made for people who are considered minimally verbal or nonverbal has become apparent. As a result, there has been a push for research on assessments in this population which has included the development of the Communication Complexity Scale (CCS); the CCS is copyrighted by the University of Kansas (see Appendix).

In 2012, Brady and colleagues established a new observational tool, the CCS, which was meant to address the need for collecting standardized data on individuals who primarily communicate without speech. The CCS is an independent scale that can be applied to various assessment or observational contexts. It was developed to describe communication of increasing degrees of coordination between a referent and communication partner until markedly sophisticated communication is used, such as verbal speech. Brady et al. created the CCS due to limited available assessments designed to measure and differentiate communication in people who primarily use pre-symbolic or early symbolic communication. Prior to the development of the CCS, most available assessments relied on caregiver reports and were not sensitive to subtle behavior changes in various populations; furthermore, they relied on communicative behaviors, such as vision, that may be difficult for individuals with multiple disabilities or disabilities such as severe blindness.

Brady et al. (2012) developed an ordinal scale with 11 different levels; authors from three universities contributed to the development, and initial adjustments were made for the inclusion of various populations. Participants included preschool-age children with intellectual and developmental disabilities, infants between ten and 36 months of age with moderate-to-severe

motor impairments who were nonverbal and not producing intentional communication signals, and individuals with severe and multiple disabilities between the ages of seven and 60 with suspected vision impairments.

Initial interpretations of the scale revealed high interrater reliability: .98 for the total sample, which consisted of 92 applications of the scale; furthermore, 92% of the assessments fell within one point of each other. The CCS was compared with several other standardized measurements and informant report measures, including the expressive language scale of the Mullen Scales of Early Learning (MSEL), the expressive scale of the Preschool Language Scale, Fourth Edition (PLS-4), the Communication Matrix, and the Communication and Symbolic Behavior Scales Developmental Profile (CSBS). It was found that the Pearson correlation was significant at the .01 level between the MSEL Expressive Language scale raw score and the CCS, as well as the raw score of the PLS-4 Expressive scale and CCS. The CCS and Communication Matrix scores were similar but not identical when comparing symbolic, intentional, and pre-intentional scores across 15 participants. However, the Spearman rho correlation for total raw score of the CSBS questionnaire and CCS was insignificant with a p-value of .94. Therefore, the CCS better correlated with the MSEL Expressive Language scale, PLS-4, and Communication Matrix in comparison to the CSBS.

Several expert opinions were solicited as a test of validity as well; results of these interviews suggested that there is a need for the CCS scale, but adjustments were needed. The authors concluded that the CCS is applicable across a range of populations and ages, and the initial development of the scale reflected an expected range in performance for various participants. CCS scores were found to be correlated with other similar available measures but

not redundant. Further research was recognized as necessary for further development of the scale, specifically the upper end, which reflects largely symbolic scores.

In 2018, Brady et. al examined the reliability and validity of a revised version of the CCS. This new, revised version of the CCS adjusted the scale from an ordinal scale with 11 potential scores to an ordinal scale with 12. Additionally, the possibility to score communication function was added for scores that are considered intentional, whether pre-symbolic or symbolic. The function options include joint attention (JA) for commenting, and behavior regulation (BR) for requesting. Two separate studies of varying populations were conducted as part of the overall project: both required participants to have minimally verbal skills, defined as producing less than 20 functional words. The first study examined the inter-observer agreement, test-retest reliability, and concurrent validity of 239 participants with intellectual disabilities between the ages of three and 66. CCS scores were compared with scores from the Vineland Adaptive Behavior Scales, Second Edition, and the Communication Matrix. The second study examined the reliability and concurrent validity of the CCS scores from children (between ages three and nine) with autism. CCS scores for this study were compared with scores from the Early Social Communication Scales.

The CCS has been used in several studies as a form of assessment. Thiemann-Bourque, Brady, and Hoffman (2019) used the CCS scale to measure changes in communication of 23 preschoolers with a diagnosis of autism. The CCS was used to code both a pre-assessment and post-assessment of each participant, who were randomly assigned to either a peer-mediated approach that incorporated speech-generating-devices (SGD) or a control group with untrained peers, considered “business as usual.” The CCS was used to score two structured contexts for each student (one with an adult, one with a peer partner), each which lasted 30 minutes. The

results indicated that children in both groups showed significant changes in their CCS scores between pre-treatment and post-treatment. The results of this study indicated that the CCS was sensitive to change over time; however, it did not demonstrate if the CCS was sensitive to change due to treatment versus maturation. This study demonstrated preliminary support that the scale can be used to measure communication changes in various contexts.

The CCS has also been used in several studies as one of several assessment measures while investigating participants with limited expressive language skills. Brady, Thiemann-Bourque, Fleming and Matthews (2013) investigated a model of language development for nonverbal preschool-age children who were learning to communicate through augmentative and alternative communication (AAC). The model suggested that children's intrinsic predictor of language is a latent variable consisting of cognitive development, comprehension, play, and nonverbal communication complexity; additionally, mediators of vocabulary acquisition included adult input at home or school, and the amount of augmentative and alternative communication. The project used a longitudinal, within-subjects design in which participants were tested at two separate times, between 10.4 and 17.6 months apart; the CCS was used to measure the participant's pre-symbolic and early symbolic communication through a scripted communication assessment, among a battery of other assessment measures, including the MSEL and PLS-4.

In 2015, Brady and colleagues investigated an intervention package with the target to increase expressive word learning in school-age children with autism who had limited expressive vocabularies. Ten children between six and 10 years of age participated in the study, which used speech sound practice and AAC to teach individualized vocabulary words. Among the assessments used in the study were the Vineland Adaptive Behavior Scales, Second Edition,

Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4), imitation through the Early Steps Imitative Sequences Assessment, as well as the CCS. In the results of the study, participants who responded well to intervention had a higher baseline for both their PPVT-4 and their CCS.

Live scoring of assessments

In each of the previously mentioned studies, the CCS was applied to structured play assessments which were video-recorded; the CCS was used to score assessments at a later point in time from the original observation, rather than during the assessment itself. Therefore, the validation of the CCS relied heavily on the ability for researchers to watch an assessment and apply the score at a delayed time. Clinically speaking, the ability to score assessments live is valuable, and many currently valid evaluation methods exist in which live analysis of a participant is used.

In the realm of education, there has been some research dedicated to live and video recorded measures in the classroom. In 2016, Curby and colleagues compared live and video recorded observations of the Classroom Assessment Scoring System (CLASS-PreK). The CLASS-PreK is used to measure the quality of interactions that teachers have with students. In this study, it was used in comparison to several child outcome measurements, including the PPVT-4, Test of Early Math Ability, Third Edition (TEMA-3), and Test of Early Preschool Literacy (TOPEL) to predict correlation between the child measurements with the CLASS-PreK. The researchers found there were generally high correlations between the live coded and video coded versions, although video scores were slightly lower on average than live scores. Cronbach's alpha was calculated for various indicators established in the CLASS-PreK to determine internal consistency reliability across live and video assessments. The results suggest that video is not the preferred medium for understanding the aspects of setting, in this case a

classroom, reflected in an alpha coefficient under .70 for negative environment. However, in comparison with the child outcome measurements, it was found that video and live coding illustrated differential prediction to student academic success when examining emotional support, classroom organization, and instructional support, suggesting that live scoring captures different or additional aspects in comparison to live coding.

In the field of speech-language pathology, dependence on a clinician's ability to reliably assess human behavior through observation is a necessary tool for data collection and progress monitoring in the profession. Researchers have explored the theory behind establishing reliable clinical observers, ultimately concluding that observations can be considered reliable if differences in recorders or observational contexts do not affect data and the conclusions drawn from them (Cordes & Ingahm, 1994).

While examining the reliability and validity of recorded versus live language measurement, researchers have considered a variety of contexts. The widespread practice of sampling for language sample analysis has been evaluated to compare live and video results. In 1991, Klee, Membrino, and May studied the differences between orthographically transcribing language samples live on computers versus through transcription done from audio tapes. Participants in the study were preschool-aged children who were evaluated either for initial services or for retesting after a period of intervention. Two speech-language pathologists served as transcribers, each transcribing half of the samples in real-time and half of the samples through audiotape. It was found that the live analysis captured approximately 90% of the sample, although no significant differences existed between the real-time orthographic samples and those written from audio tapes.

Later research by Furey and Watkins (2010), studied the ability to capture specific aspects of language production in a similar population group. This study focused on real-time language analysis of verbs in comparison to recorded analysis of the same word-class. Participants included two groups, half of which were typically-developing preschoolers and half of which were preschoolers with a language impairment. Results demonstrated a strong, positive correlation between the two mediums of data collection (overall correlation between live versus audio was between $r = .82$ to $r = .90$); however, limitations of the study include that it only focused on elicitation of verbs.

Not only has the reliability of language sample analysis between live and recorded collection methods been studied, but the clinical preferences of speech-language pathologists have also been considered. A 2016 nationwide survey of 1,399 practicing speech-language pathologists in the school setting revealed that over half of SLPs did not use any type of recording (audio or video) when conducting language samples, and rather, transcribed the sample as the child spoke (Pavelko, et al., 2016).

Research has also compared accuracy for coding disfluencies using live versus video coded scores. Speech-language pathologists in the field traditionally used various methods to analyze a speaker's speech through recording. In 1998, Yaruss established a specific method for real-time fluency transcription which meant for quick, easy, and clinically applicable collection of data. Yaruss and colleagues (1998) set out to compare real-time and transcript-based measurements of stuttering, effectively comparing a live versus recorded version of participant's speech. Using 50 audio-taped recordings, both a transcription method and real-time analysis method were used to study the disfluencies of the speakers. Clinician's using the transcription-based method collected a 200-syllable speech sample which was then orthographically

transcribed from an audio recording and analyzed for various types of disfluencies. In contrast, clinicians using the real-time method either listened live or in real-time to the recoding of a person's speech, and simply marked the number of disfluencies in the total number of words. Types of disfluencies were considered either more typical (disfluencies characteristic of individuals who do not stutter) or less typical (characteristic of individuals who do stutter). It was found that there was consistency between the two analyzation methods, with some predictable differences, notably brief, subtle prolongations or multi-component, clustered stuttering events. Additionally, using a severity rating of four possible levels, language samples were reliable between coding methods, only varying twice among the 50 participants. It was established that both methods were considered reliable, in which the transcription approach was suggested for use when more detailed information is needed while the real-time method may be used to track more general progress.

Participant and assessment variables

Various factors may influence the ease of administering and scoring various assessments, including participant and assessment variables. In this study, three participant variables are examined for their potential influence on live scoring the CCS in comparison with video scoring: communication level, age, and severity of autism. Additionally, the environmental variable of type of assessment task (joint attention or behavior regulation) was analyzed.

Communication level

In the field of speech-language pathology, assessments are frequently centered on language rather than overall communication. Research participants that are nonverbal communicators require additional considerations for proper and accurate assessment. In 2013, Kasari and colleagues examined best practice measures required for the assessment of school-

aged children who are minimally verbal with a diagnosis of autism spectrum disorder. It was found that the assessment needs of minimally verbal children with autism had been overlooked, with many of the available assessments for language and social skills having serious limitations in their application to this population. Kasari and colleagues concluded that these assessment limitations contribute to the lack of information about this population. Given that children with autism experience a range in severity of various associated behaviors, including sensory preferences or difficulty with change in routine, it is necessary to consider how a combination of these traits, measured through overall severity, may affect the ability for a child's communication to be evaluated effectively. Distinguishing between self-regulating behaviors and communication directed towards others can often be a difficult task, conflated by a variety of factors including the ability of an evaluator to collect additional information through video coding in comparison to live coding.

Age

Autism is a developmental diagnosis which influences a person across their lifespan. Research continues to grow in the field of studying the effectiveness of early intervention for children with autism. The age of diagnosis has continued to lower, with support for early intervention to further develop better child outcomes later in life. However, the supports used during early intervention are likely to change once a child enters school, and as such, research is needed in assessment and treatment practices to support all phases of development (Landa, 2018). Increased diagnosis of autism in recent years has called on researchers to examine psychosocial interventions for individuals across the lifespan (Damiano et al., 2014). There is an established need for research on individuals with autism of all ages, with special consideration to how age may affect assessment and treatment needs.

Severity of autism

The current Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) definition of autism includes a reference to the severity of autism, which considers levels of need as well as language, cognitive, behavioral and adaptive skills. The DSM-5 references levels of support needs as the defining factor in severity, with reference to needs in both communication and behavior. Mehling and Tassé (2016) conducted a literature review to determine the current best assessment sources for professionals to use in establishing the severity of autism in their clients. Results of the literature review suggested that the current most common sources for determining severity of autism include the Autism Diagnostic Observation Schedule (ADOS), Autism Diagnostic Interview, Revised (ADI-R), Child Autism Rating Scale (CARS), and Social Responsiveness Scale (SRS). The DSM-5 criteria focus on the person-environment influence on level of needs, rather than singular focus on within-person deficits. Fuller understanding through formal measurement and clinical judgment of a child's severity of autism may inform the best environment for assessment, thereby influencing the medium of the assessment (live versus video-recording).

Communication function

Joint attention is considered sharing an experience or an object with another person through vocalizations and shifting eye-gaze. In typically-developing infants, the ability to respond to acts of joint attention demonstrates important social, cognitive and regulatory skills for the development of language. It is first seen at approximately nine months of age with parents, and later with peers (Delgado, et al., 2002). However, in individuals with ASD, the ability to engage in joint attention is impaired, with difficulty sharing and lack of engagement with peers both part of the DSM-5 criteria for a diagnosis of autism. Although children with

autism may need intervention to target a variety of communication needs, such as requesting or asking questions, their specific difficulty with social engagement makes it crucial to understanding its influence on a person with autism's ability to communicate (Kourassanis-Velasquez & Jones, 2018).

Purpose of the current project

The purpose of this project is to reduce the time-cost of the video application of the CCS. Professionals have limited time left for documentation, writing, research, and more; employers frequently expect service-care providers to spend a majority of their time interacting with clients to maximize their productivity, leaving limited instances in which a professional can spend time coding and analyzing data collected through assessments such as the CCS. Additionally, assessments conducted in other areas in the field suggest real-time analysis is often used by professionals (Pavelko, et al., 2016) and strongly correlated with recorded methods of evaluation (Furey & Watkins, 2010; Klee et al., 1991; Yaruss et al., 1998). Through elimination of the post-assessment coding portion of the CCS, the proposed project serves to provide a valuable tool for professionals who are interested in the CCS for its applicability to the minimally verbal population, but currently cannot use it due to time-constraints.

Educators work with a variety of students who communicate nonverbally, and as a result these students have a variety of needs when they are working one-on-one with an adult. Some children may potentially be excited about an assessment, engaged in toys, communicate in a clear manner, and have a limited number of distracting behaviors. These children are the optimal candidate for live coding of the CCS. Conversely, other children may find it difficult to sit for long periods of time, have complex medical or behavioral needs, or communicate in a subtle or more difficult-to-interpret style. These children may benefit from the additional time taken to

score the CCS through video. Ultimately, given the availability of recording equipment at disposal in modern-day school and clinical settings, it is recommended that a video of an assessment for analysis with the CCS is always obtained. This ensures that regardless of if an assessment can be scored accurately by the CCS live or not, the educator has a copy of the assessment at their disposal for additional information or coding as needed. By seeking to validate live coding of the CCS, this project seeks to provide well-informed additional measures for persons who are nonverbal, including children with autism; live coding of the CCS is meant to be a complimentary method of the traditional video coding of the CCS, not a complete replacement.

Research questions

There are four research questions to be addressed. The first research question is, Does the effect of type of coder (live and video) on optimal/typical communication differ based on the communication level (pre-intentional non-symbolic, intentional non-symbolic, or intentional symbolic) of the child being assessed? The hypothesis is that there will be a significant difference between type of coder when the child being assessed scores in the intentional non-symbolic or pre-intentional non-symbolic category of communication. The second research question is, Does the effect of type of coder (live and video) on optimal/typical communication differ based on the age of the child being assessed? The hypothesis is that there will be a significant difference based on the age of the child being assessed. The third research question is, Does the effect of type of coder on optimal/typical communication differ based on the severity of autism of the child being assessed? The hypothesis is that there will be a significant difference based on the severity of autism of the child being assessed. The final question is, Does the effect of type of coder on individual task score differ based on the function (joint attention or behavior

regulation) of the task being assessed? The hypothesis is that there will be a significant difference based on the function of the task being assessed.

Chapter 2: Methodology

In order to answer the research questions, research and data collection was gathered through the research project *Measuring early communication in students with autism* (MECSA).

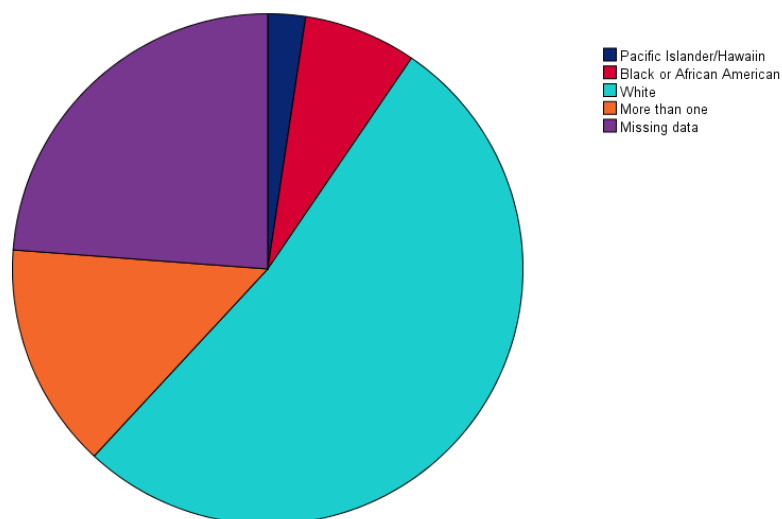
Participants

Demographics

The participants in the study included school-aged children (three to 18 years) with severe autism. The total number of children who completed an assessment in which the CCS was coded both live and through video was 42. There were 33 males and nine females enrolled in the study. The racial demographics include: 22 white, three Black or African American, one Hawaiian/Pacific islander, and six more than one race; ten participants had missing data for racial background, as seen in Figure 1. Inclusion criteria were: educational or medical diagnosis of autism using current DSM-5 criteria, and “minimal verbal” skills at the time of their participation in the study. Minimal verbal was defined as less than 40 different words in their total vocabulary, confirmed by the child’s special educator as well as researchers in the project. The children were recruited from public schools of the school districts in Lawrence, Kansas and the greater suburban area of Kansas City, Kansas. The school districts included were Lawrence, Shawnee Mission, Olathe, and Blue Valley. Children were referred to the study through their school educators, typically either special education teachers or speech-language pathologists. Assent to participate was obtained through informed consent by the children’s caregivers. Observational assessments were completed at the child’s school during the school day, either within the typical school year or during extended school year over the summer.

Figure 1

Race and ethnic demographics of participants



Measures

Vineland Adaptive Behavior Scales

The Vineland Adaptive Behavior Scales, Third Edition, is a behavior rating assessment that uses caregivers or teacher report to aid in diagnosis and development of treatment plans for individuals with intellectual and developmental disabilities. The assessment asks caregivers or teachers to rate individuals in the communication domain (expressive, receptive, written), daily living skills domain (personal, domestic, community), and socialization domain (interpersonal relationships, play and leisure, coping skills). It was normed on a sample representative of the United States population, which was controlled for age, race, gender, geographic region, socioeconomic status (SES), education of parents, and with inclusion of special populations such as autism. Subjects in this study were administered the Vineland-3 via phone call with a parent or caregiver; caregivers were compensated for their participation with a \$50 gift card and received a summary of official results of the Vineland-3 via a mailed results letter. Parents were offered access to the full Vineland-3 report upon request. Of the 42 participants included in this

project, 30 participants received a Vineland-3 administration; all 42 participant's caregivers were offered the chance to complete the caregiver report. One participant submitted a recent copy of the Vineland-3 that had been administered for a school evaluation within several months of the planned assessment date for the research project. Results of the Vineland-3 suggested that most study participants scored below the first percentile in both their overall adaptive behavior composite score and the communication domain, although several outliers existed outside of this trend. The adaptive behavior composite standard scores ranged from 28 to 107, with an average of 55.6; communication domain standard scores ranged between 20 to 103, with an average of 44.9. Additionally, most of the participants scored the communication domain as an individual weakness. In cases where the communication domain was not an individual weakness, the majority scored receptive language or written language as strengths, rather than expressive language.

Childhood Autism Rating Scale

The Childhood Autism Rating Scale, Second Addition (CARS-2) is an observational rating scale meant to identify children with autism as well as determine autism severity through quantifiable ratings assigned by direct observation. It is meant for individuals ages two and older, and on average takes up to ten minutes to score if there has been adequate observation of the individual. The assessment includes 15 items, each with a possible of seven scores to assign the child and specific behaviors or qualifications that would make certain scores appropriate. The CARS-2 assessment was administered by staff of the MECSA project to each student participant; a minimum cutoff score of 30 out of 60 was defined for participation in the project. Analysis of the assessments show that the range of the participant's CARS-2 scores are between 31.5 and 48.5.

Application of the CCS

The CCS was used to score a structured play-assessment consisting of 12 “sabotaged” games that were manipulated to promote communication. In both the live and video versions of the assessment, scores were assigned to each of the 12 tasks. Six of the tasks were designed to elicit joint attention, while six of the tasks were designed to elicit behavior regulation. Two types of descriptive summary scores were calculated for each assessment using the scores from each of the 12 tasks: optimal scores (average of the first quartile, or an average of the highest three scores out of twelve) and typical scores (average of the second and third quartiles, or an average of the middle six scores out of twelve). For the purposes of this project, joint attention is defined as a communicative function that is used to direct the communication partner’s attention to an object, event, or activity by commenting, gesturing, or vocalizing, and the desired outcome is shared interaction. Conversely, behavior regulation is defined as a communicative function used to obtain a specific result by either requesting or protesting an object, action, or activity. Examples of tasks used in the structured play assessment meant to elicit joint attention include the discovery of misplaced items in various toys (ex. a bug in a set of blocks), playing music, or the unexpected activation of a light fan. Examples of tasks used in the structured play assessment meant to elicit behavior regulation include toys that are broken (ex. a hammer toy with no batteries) or a snack that is difficult to access because the container is sealed shut. Thirty-nine of the 42 participants in the assessment used a child’s version of the assessment, while three of the participants used a similar set that was created with age-appropriate tasks for teens or adults.

The application of the CCS during the live administration of the structured play assessment was then compared with the application of the CCS to a delayed, video-coded assessment. It was a within-subjects design, as each assessment administered received both a set

of live scores and a set of video scores. Standard analysis used included examining scores to determine if the same coding of behavior was given to the child for each of the 12 tasks, as well as analysis for optimal scores and typical scores. For analysis purposes, it was assumed that video-coded assessments are most accurate, with live scores compared against their video pair. See the appendix for full description of the CCS scale and scoring sheets.

Coder criteria

Criteria for CCS coders was established to determine their inclusion in the project. A coder must have had reliably scored a video using the CCS with 75% accuracy in three out of five consecutive videos. Additionally, the coder must have been within one numerical value of the optimal score as well as the typical score, defined above. Past data suggests that this standard of reliability takes a range of 18 to 35 videos to meet.

To be a reliable live coder, the person must have met the standards listed for a video coder. Additionally, the coder must have been considered reliable at applying the CCS scale to naturalistic observations. To be a reliable live coder, the person must have scored three naturalistic observations in the classroom with 75% accuracy in comparison with an established, reliable naturalistic coder. In comparison to a reliable coder, the person must also have been within one point of the optimal score and one point of the typical score. The coder must have practiced becoming reliable on more than one child, and on multiple different school days. These standards were determined based on current video-scoring and naturalistic reliability standards in the MECSA project.

There was a total of 11 people who met these standards of inclusion for video-scoring, and ten individuals who met the standard for naturalistic scoring using the CCS. Of the potential 11 coders qualified for the project, ten participated as both video coders and live coders, and one

participated as a video coder only. Ten of the coders were graduate research assistants in the departments of either speech-language pathology or psychology; one was the project coordinator.

General analysis approach

The research questions of this study focus on comparing live application of the CCS with video application of the CCS, when considering the participant variables of communication level, participant age, and severity of autism, as well as the assessment variable, function of scripted assessment task. In order to answer these questions, several methodological approaches were used. In order to examine the participant variables (communication level, age, and severity of autism), the calculation of summary scores was employed. This included comparison of optimal and typical scores based on coder type. For the assessment variable, individual task-by-task analysis was used as a direct comparison when considering coder type.

Two main methods of analyses provided information for the research questions addressed: statistical modeling and descriptive data analysis. To examine each of the participant variables using statistical modeling, a repeated measures general linear model was calculated to examine within-subjects' effects and determine if a variable by coder type interaction existed. In order to examine the assessment variable, a series of independent t-tests were calculated for each task.

Descriptive data analysis was used to gain a fuller understanding of the relationships between coder type and the variables assessed. Methods included calculation of the percent agreement between optimal and typical scores, both as overall scores as well as after the optimal and typical scores were separated into groups based on the video communication level. This gave information on how many live-video scripted assessment pairs fell within one point of each other for both optimal and typical scores overall, as well as by communication level. Additionally, in

studying the function of scripted assessment task, the average differences between task-by-task scores were calculated to determine if there was greater variability depending on task type.

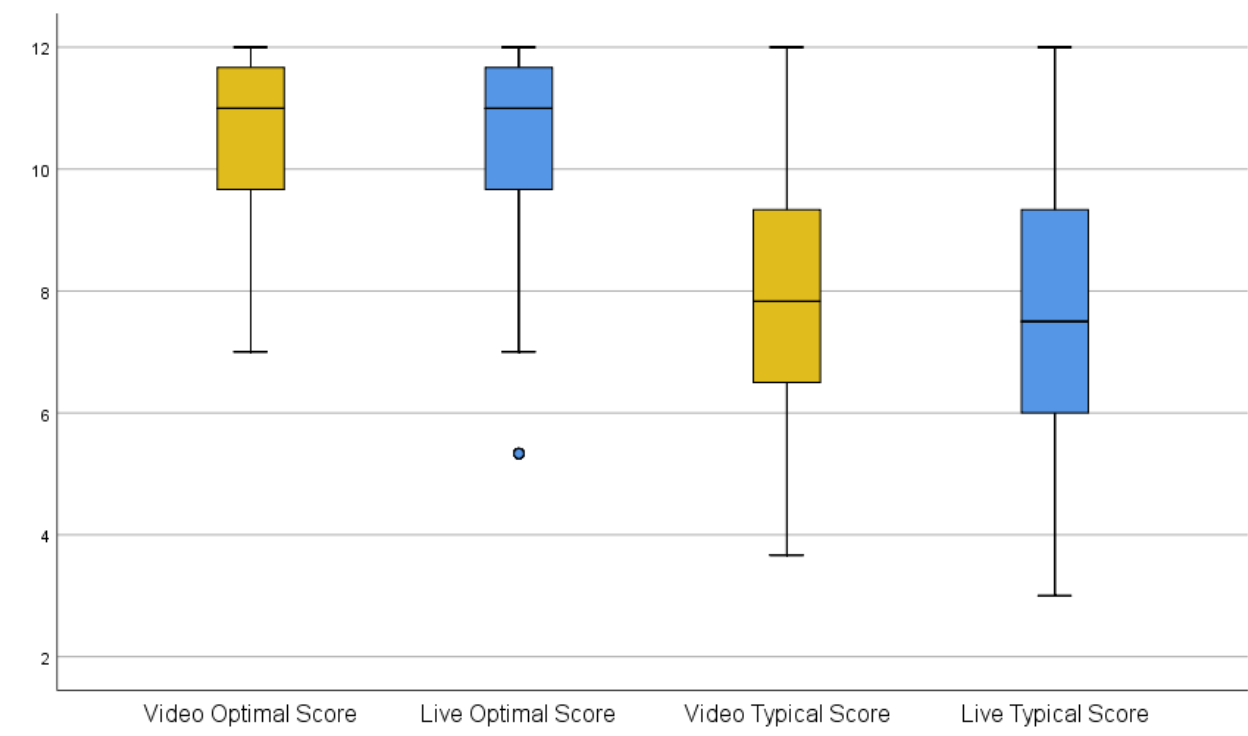
Chapter 3: Results

Preliminary descriptive information

To lay the foundation for answering later research questions, an examination of overall descriptive measures of the CCS optimal and typical scores was considered. Figure 2, below, represents the overall average of optimal and typical scores by coder type for the 74 assessments collected during the study. As expected, optimal scores for both coder types are much higher with a smaller range of scores (about seven to 12), while typical scores for both coder types are on average lower and with a much larger range of scores (about three to 12). These results not only represent the average scores for each type of summary score, but also reflect a greater diversity of scores overall when comparing typical scores to optimal scores.

Figure 2

Distribution of optimal and typical scores in video and live coders



To examine initial agreement levels, differences between live and video coded scores of both the optimal and typical scores were calculated. The frequency of differences was summarized to indicate what percent were within perfect agreement (within zero points) and what percent were within one point, as seen in Table 1. Overall agreement was higher for optimal scores when considering percent agreement of within zero points as well as agreement within one point.

Table 1

Difference in live versus video optimal and typical scores

	% agreement within 0 points	% agreement within 1 point
Optimal Score Difference	52.7	83.8
Typical Score Difference	14.9	66.2

Research Question 1: Communication level of participant

Does the effect of type of coder (live and video) on optimal/typical communication differ based on the communication level (pre-intentional non-symbolic, intentional non-symbolic, or intentional symbolic) of the child being assessed? Two types of analysis were performed to answer this question: a repeated measures general linear model as well as descriptive measures of frequency.

In order to determine if there is significant difference between the live and video typical and optimal scores based on communication level, a within-subjects design was used in which the score by the live coder and score by the video coder was considered the dependent variable while the optimal or typical video communication level was considered the independent variable, with the following three levels. Communication levels were defined for either the optimal or the typical score for both the live and video scores through the following categories: scores of zero

to five were defined as pre-intentional non-symbolic, scores of six to ten were defined as intentional non-symbolic, and scores of 11 and 12 were defined as intentional symbolic. Pre-intentional non-symbolic scores included behaviors such as single orientation to the task or assessment administrator, single orientation in combination with some vocalizations and/or gestures, and dual orientation to the task or assessment administrator without additional behaviors. Intentional non-symbolic scores included dual orientation to the task and assessment administrator in combination with vocalizations and/or gestures, triadic orientation between the task and assessment administrator, and triadic orientation between the task and assessment administrator in combination with gestures and/or vocalizations. Intentional symbolic communication included language, either words and phrases, communicated to the assessment administrator through speech, a speech-generating device, a picture exchange system (PECS), or sign language.

Analysis for the difference between live and video optimal score based on communication level revealed that there is not a significant difference ($F_{1,72} = .67, p = .42$). Analysis for the difference between live and video typical score based on communication level revealed that there is not a significant difference ($F_{1,72} = .13, p = .88$). See Tables 2 and 3 for full descriptive statistics regarding live and video optimal and typical score based on communication level. For both optimal and typical scores, the difference in the score means were within one point regardless of communication level of the participant.

Table 2

Optimal score communication level descriptive statistics

Communication Level	Coder Type	Score Mean	Std. Deviation	N
Intentional Non-Symbolic	Video	9.11	1.00	32
	Live	9.25	1.33	32
Intentional Symbolic	Video	11.46	.43	42
	Live	11.47	.56	42

Table 3

Typical score communication level descriptive statistics

Communication Level	Coder Type	Score Mean	Std. Deviation	N
Pre-intentional Non-Symbolic	Video	4.72	.64	13
	Live	4.63	.97	13
Intentional Non-Symbolic	Video	8.06	1.28	53
	Live	7.81	1.59	53
Intentional Symbolic	Video	11.50	.45	8
	Live	11.23	.75	8

Descriptive frequencies were calculated to measure the difference in type of coder scores within the communication level. According to video coded scores, of the optimal communication scores, 32 optimal scores fall within the intentional non-symbolic communication level while 42 optimal scores fall within the intentional symbolic communication level. Of the participants assessed, there were no optimal scores within the pre-intentional non-symbolic communication level. Of the assessments within the intentional non-symbolic communication level, 71.9% of the optimal scores have coder agreement within one point. Of the assessments within the intentional symbolic communication level, 92.9% of the optimal scores have coder agreement within one

point. According to video coded scores, of the typical communication scores, 13 typical scores fall within pre-intentional non-symbolic communication level, 53 typical scores fall within the intentional non-symbolic communication level, and 8 typical scores fall within the intentional symbolic communication level. Of the assessments within the pre-intentional non-symbolic communication level, 69.2% of the typical scores have coder agreement within one point. Of the assessments within the intentional non-symbolic communication level, 62.3% of the typical scores have coder agreement within one point. Finally, of the intentional symbolic communication level, 87.5% of the typical scores have coder agreement within one point. According to descriptive data, the intentional symbolic communication level has overall higher video and live coder agreement for both optimal and typical CCS scores, in comparison to the pre-intentional non-symbolic and intentional non-symbolic communication levels.

Research Question 2: Age of participant

Does the effect of coder type (live and video) differ based on the age of the participant, as measured through optimal and typical score? To determine if there is an effect, a repeated measures general linear model was calculated. A within-subjects design was used, by which coder type (live or video) was considered the dependent variable while the optimal or typical video score was considered the independent variable, and age of participant was the between-subject's predictor of interest. Analysis for the difference between live and video optimal scores based on age at administration revealed that there is not a significant age by coder type interaction ($F_{1, 71} = .19, p = .66$). Analysis for the difference between live and video typical scores based on age at administration revealed that there is not a significant age by coder type interaction ($F_{1, 71} = .44, p = .51$).

Research Question 3: Severity of autism of participant

Does the effect of coder type (live and video) on optimal and typical scores differ based on the severity of autism of the participant? To determine if there is an effect, a repeated measures general linear model was calculated. A within-subjects design was used, by which coder type (live or video) was considered the dependent variable while the optimal or typical video score was considered the independent variable, and the severity of autism of the participant was the between-subject's predictor of interest. Analysis for the difference between live and video optimal scores based on severity of autism revealed that there is not a significant severity of autism by coder type interaction ($F_{1,69}=.36, p=.55$). Analysis for the difference between live and video typical scores based on severity of autism revealed that there was not a significant severity of autism by coder type interaction ($F_{1,69}=1.26, p=.27$).

Research Question 4: Function of scripted assessment task

Does the effect of type of coder on individual task score differ based on the function (joint attention or behavior regulation) of the task being assessed? A comparison of the means between the live coder and the video coder was conducted for each task; the difference was calculated by subtracting the live coder scores from the video coder scores. The three participants who were administered the adult-level assessment were excluded from the analysis, given the small size of the group. It was found that average video coder scores were nominally higher than the average live scores in all but four tasks, where it was found that the average live scores were higher. The greatest mean difference in coder scores for a task was .41, with the total average difference of .10 with a standard deviation of 1.95. Three of the four tasks with higher live than video scores had the function of behavior regulation (BR), while just one task had the function of joint attention (JA). See Table 4 for full descriptive differences between live and

video means. This suggests neither group of coders score more conservatively on one type of task-function than the other.

Table 4

Comparison of the difference of task score means between live and video coders

Task Name	Difference of means	N	Std. Deviation
Wind Up (BR)	-.17	70	1.79
Blocks (JA)	.09	69	2.19
Snack (BR)	-.03	71	1.41
Music (JA)	.37	71	2.06
Hammer (BR)	.15	71	1.92
Fan (JA)	-.04	70	1.96
Magna Tiles (BR)	-.11	71	2.68
Dots (JA)	.14	69	2.50
Bubbles (BR)	.07	71	1.59
Books (JA)	.11	71	1.39
Bumble Balls (BR)	.17	69	1.64
Ball Toy (JA)	.41	68	2.23
Total	.10	877	1.95

A paired samples t-test was performed in which the video score and live score were paired for the twelve tasks. Effect size was calculated for all tasks, which revealed that all twelve tasks had a small effect size, suggesting that the difference between the live and video coder scores were not significant regardless of task type. An independent t-test was calculated in which the scripted assessment tasks were grouped into their category function: behavior regulation or joint attention. It was found that for the function behavior regulation, the average difference between live and video coders is .01 with a standard deviation of 1.88; for the function joint attention, the average difference between coders is .18 with a standard deviation of

2.07. There is not significant differences between functions and the effect size is very small, $t(828.91) = -1.21, p = .23, d = .08$. See Table 5 for a description of task score means, and effect sizes for each task.

Table 5

Task score means and effect sizes by task

Task Name	Coder type	Mean	N	SD	Paired samples <i>t</i> -test, Cohen's <i>d</i>
Wind Up (BR)	Video	7.11	70	3.24	$t(69) = -.80, p = .43, d = .10$
	Live	7.29		3.34	
Blocks (JA)	Video	7.01	69	3.46	$t(68) = .33, p = .72, d = .04$
	Live	6.93		3.49	
Snack (BR)	Video	8.92	71	2.75	$t(70) = -.17, p = .87, d = .02$
	Live	8.94		2.72	
Music (JA)	Video	7.45	71	3.26	$t(70) = 1.50, p = .14, d = .18$
	Live	7.08		3.17	
Hammer (BR)	Video	7.41	71	2.84	$t(70) = .68, p = .50, d = .08$
	Live	7.25		2.88	
Fan (JA)	Video	7.04	70	3.58	$t(69) = -.18, p = .86, d = .02$
	Live	7.09		3.56	
Magna Tiles (BR)	Video	7.30	71	3.07	$t(70) = -.36, p = .72, d = .04$
	Live	7.41		3.16	
Dots (JA)	Video	7.59	69	3.15	$t(68) = .43, p = .63, d = .06$
	Live	7.45		3.40	
Bubbles (BR)	Video	9.04	71	2.83	$t(70) = .37, p = .710, d = .04$
	Live	8.97		2.96	
Books (JA)	Video	8.04	71	3.02	$t(70) = .68, p = .50, d = .08$
	Live	7.93		3.11	
Bumble Balls (BR)	Video	7.86	69	3.10	$t(68) = .88, p = .38, d = .11$
	Live	7.68		3.26	
Ball Toy (JA)	Video	7.50	68	2.90	$t(67) = 1.52, p = .13, d = .11$
	Live	7.09		3.07	

Chapter 4: Discussion

This study examined the reliability and validity of live coding of the Communication Complexity Scale (CCS) in comparison with the established video coding of the assessment (Brady et al., 2012; Brady et al., 2018). In order to determine possible factors in the reliability of live and video scoring, this study examined the effects of participant communication level, participant age, participant severity of autism, and the designed communication function of the tasks in the assessment (joint attention or behavior regulation). Overall levels of agreement for the CCS when comparing live and video scores demonstrated that agreement regarding optimal scores were generally higher than agreement for typical scores.

Communication level of participant

Overall analysis between communication levels for both optimal and typical communication scores found that there was no significant difference between live and video coders when examining the effect of participant communication levels (pre-intentional non-symbolic, intentional non-symbolic, or intentional symbolic).

Further descriptive analysis found that there were differences in the percentage of live coder and video coder pairs who had a difference of optimal/typical score by more than one point. It was found that for participants with an intentional non-symbolic optimal score, 71.9% of coders had a difference of less than one point, while for participants with an intentional symbolic optimal score, 92.9% of coders had a difference of one point. No participants in the study had an optimal score in the pre-intentional non-symbolic communication level. Results are consistent with the hypothesis that it would be more difficult for coders to be consistent when scoring participants whose highest communication acts are in the intentional non-symbolic communication level. This is likely because the intentional non-symbolic communication level

relies on the use of dual and triadic orientation, characterized by touch and eye-gaze, often paired anecdotally with gestures and vocalizations. This is typically considered more difficult for coders to reliably score in comparison with intentional symbolic communication, which consists of words and phrases.

These results for optimal scores are consistent with the descriptive analysis conducted for typical scores of participants. As with the optimal score, analysis found differences in the percentage of live coder and video coder pairs who had a difference of typical score by more than one point based on communication level. Of the participants that most typically communicated at the pre-intentional non-symbolic level, 69.2% of the typical score's had coder agreement within one point, while participants that typically communicated at the intentional non-symbolic level, 62.3% of the typical score's had coder agreement within one point. This contrasts with the participants who most typically communicated at intentional symbolic level, in which 87.5% of the typical score's had coder agreement within one point. These results suggest that it is more difficult for coders to reliably assign scores to participants when they typically communicate at the pre-intentional non-symbolic level or intentional non-symbolic level, likely because these categories are both made up of gestures, vocalizations, eye-gaze, and touch with varying degrees of combination and complexity. In comparison to symbolic intentional communication, which is comprised of language, pre-intentional non-symbolic and intentional non-symbolic communication acts are subtler and require more training to understand given that they are considered "nonverbal" communication and registered less overtly in comparison to language for many people.

Age of participant

Analysis regarding age of participant revealed that for both optimal and typical scores, there was no significant difference within-subjects for the reliability of scores based on type of

coder. This suggests that regardless of specific age, children with autism within the larger category of “school-age” (ages three to 18) can be reliably administered the CCS without differences in scores based on live and video coding. This ultimately demonstrates that while age may play in an important role in the effects of autism assessment and treatment, individuals who are minimally verbal communicate, as measured by the CCS, in ways that are not significantly influenced by their age but rather likely by a host of other factors.

Severity of autism of participant

It was determined that the severity of autism of a participant was not significant for either measurement of the reliability of the optimal nor typical scores between live and video coders. Severity of autism was measured through the participant’s score on the Child Autism Rating Scales, Second Edition (CARS-2), which requires a severity rating on 15 various sections. Given the participation requirements for this project, subjects generally scored on the severe end of the spectrum for questions regarding communication, while other sections regarding sensory factors, adaption to change, etc. were variable across the participants. Overall, these results suggest that regardless of any of the other salient features of autism, these features do not alter the ability for live and video coders to score reliably against each other, nor do they predict how a participant will score on communication via the CCS. These results, however, are limited to children who have severe autism.

Function of scripted assessment task

Several analyses were performed to determine whether or not the function of the task assessed, whether that is joint attention or behavior regulation, affects the reliability of CCS scores based on coder type. Descriptively, it was revealed that there was no influence of task type on score, based on the average differences between live and video scores for each task.

Additionally, it was determined that there was not a significant difference between scores based on coder type, with small effective sizes for each task. This suggests that although people with autism may have difficulty using joint attention (Kourassanis-Velasquez & Jones, 2018), using assessment tasks meant to elicit joint attention does not significantly alter the reliability of CCS coders (live or video). With application to the CCS as an observational tool, these results suggest that when applying the tool to various forms of assessment, the coder need not worry about the activity that is being assessed when considering if the assessment should be coded live or through video.

Clinical implications

The results of this research project suggest that application of the CCS during live coding is a valid method that is not affected by a client's communication level, age, or severity of autism; additionally, it is not affected by the function of the task being assessed, whether that is behavior regulation or joint attention. The implications of this study suggest that educators should be able to choose between live coding and video coding of the CCS to best fit their needs and the needs of the client.

The potential to live code the CCS provides a large amount of new opportunities for educators and clinicians alike to use the CCS with their students or clients who are nonverbal. Previously, the CCS could only be coded through video, which required professionals to set aside time to not only administer the scale in an observational setting, but also to then watch the video back later and analyze the results accurately. In contrast, live coding of the CCS allows professionals to gain immediate results to be shared and applied, effectively cutting a large portion of the time necessary to devote to the assessment.

Although live coding provides a reliable method for applying the CCS, there are several occasions where it may be best judgement to use a video coded assessment instead. Practically speaking, individuals with behavioral needs may have difficulty sitting for an assessment. In these instances, it may be easier for a clinician to set up a video camera to record the assessment in order to ensure a safe and enjoyable test for all persons involved. Additionally, if applying the CCS to a live observation of a scripted play assessment similar to the one used in this study, two professionals will be needed: one to score, and one to administer the play assessment. If an educator or clinician cannot find an additional person for administration of the assessment or scoring of the CCS, they may choose to video record the assessment and score later to independently and reliably apply the scale.

Limitations

There are several potential limitations to the project. Although all scripted assessments will have a reliable video coder, it was not expected that all scripted assessments will be coded live by a reliable coder, given that live coder standards were based on ability to apply the scale to alternate environments rather than a live play-assessment. Furthermore, the population of the study includes young children with special needs, in which behavioral problems often arose. In these instances, the live coder was often unable to code every task administered and there were missing data points as a result. Additionally, although the CCS was developed to address all people who communicate nonverbally, regardless of diagnosis or age, this study focused on children with a diagnosis of autism, limiting the conclusions to those specific groups. In this study, the CCS was applied to a structured play assessment, suggesting that validation of application to a non-structured assessment such as observation in a naturalistic environment is still needed.

Future directions

In the future, more research will likely be needed to establish the validity of live coding the scripted assessment through a well-designed, individualized research project. Furthermore, the establishment of assessment administrator coding rather than live coding may be considered to provide another option for educators in the field to make the CCS more accessible; additionally, the application of the CCS live in other contexts beyond a structured assessment may be researched. There are several potential directions for this project given the results determined in this study. Given the descriptive results of the effect of coder type on communication level, further research may be warranted to determine if there is an effect on communication level when other variables, including age at administration and CARS score, are controlled for in the analysis.

References

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: Author.
- Brady, N. C., Fleming, K., Thiemann-Bourque, K., Olswang, L., Dowden, P., Saunders, M. D., & Marquis, J. (2012). Development of the communication complexity scale. *American journal of speech-language pathology, 21*(1), 16–28. [https://doi.org/10.1044/1058-0360\(2011/10-0099\)](https://doi.org/10.1044/1058-0360(2011/10-0099))
- Brady, N. C., Thiemann-Bourque, K., Fleming, K., & Matthews, K. (2013). Predicting language outcomes for children learning augmentative and alternative communication: Child and environmental factors. *Journal of speech, language, and hearing research, JSLHR, 56*(5), 1595–1612. [https://doi.org/10.1044/1092-4388\(2013/12-0102\)](https://doi.org/10.1044/1092-4388(2013/12-0102))
- Brady, N. C., Storkel, H. L., Bushnell, P., Barker, R. M., Saunders, K., Daniels, D., & Fleming, K. (2015). Investigating a multimodal intervention for children with limited expressive vocabularies associated with autism. *American Journal of Speech-Language Pathology, 24*(3), 438–459. https://doi.org/10.1044/2015_ajslp-14-0093
- Brady, N. C., Fleming, K., Romine, R. S., Holbrook, A., Muller, K., & Kasari, C. (2018). Concurrent validity and reliability for the Communication Complexity Scale. *American Journal of Speech-Language Pathology: 27*(1), 237–246. https://doi.org/10.1044/2017_ajslp-17-0106
- Constantino, J.N. (2012) Social Responsiveness Scale, Second Edition. Torrance, CA: WPS.
- Cordes, A. K., & Ingham, R. J. (1994). The reliability of observational data: II. Issues in the identification and measurement of stuttering events. *Journal of speech and hearing research, 37*(2), 279–294.

- Curby, T. W., Johnson, P., Mashburn, A. J., & Carlis, L. (2016). Live versus video observations: Comparing the reliability and validity of two methods of assessing classroom quality. *Journal of Psychoeducational Assessment, 34*(8), 765–781. <https://doi.org/10.1177/0734282915627115>
- Damiano, C. R., Mazefsky, C. A., White, S. W., & Dichter, G. S. (2014). Future directions for research in autism spectrum disorders. *Journal of clinical child and adolescent psychology: the official journal for the Society of Clinical Child and Adolescent Psychology, American Psychological Association, Division 53, 43*(5), 828–843. <https://doi.org/10.1080/15374416.2014.945214>
- Delgado, C. E. F., Mundy, P., Crowson, M., Markus, J., Yale, M., & Schwartz, H. (2002). Responding to joint attention and language development: A comparison of target locations. *Journal of Speech, Language, and Hearing Research, 45*(4), 715-719. [https://doi.org/10.1044/1092-4388\(2002/057\)](https://doi.org/10.1044/1092-4388(2002/057))
- Dunn, L. M., & Dunn, D. M. (2007). *The Peabody Picture Vocabulary Test, Fourth Edition*. San Antonio, TX: The Psychological Corporation.
- Furey, J. E., & Watkins, R. V. (2002). Accuracy of online language sampling: A focus on verbs. *American Journal of Speech-Language Pathology, 11*(4), 434–439. [https://doi.org/10.1044/1058-0360\(2002/046\)](https://doi.org/10.1044/1058-0360(2002/046))
- Ginsburg, H.P., & Baroody, A. J. (2003). *Test of Early Mathematics Ability, Third Edition*. Austin, TX: PRO-ED, Inc.
- Kasari, C., Brady, N., Lord, C., & Tager-Flusberg, H. (2013). Assessing the minimally verbal school-aged child with autism spectrum disorder. *Autism research: official journal of the International Society for Autism Research, 6*(6), 479–493. <https://doi.org/10.1002/aur.1334>

- Klee, T., Membrino, I., & May, S. (1991). Feasibility of real-time transcription in the clinical setting. *Child Language Teaching and Therapy*, 7(1), 27–40.
<https://doi.org/10.1177/026565909100700102>
- Kourassanis-Velasquez, J., & Jones, E. A. (2018). Increasing joint attention in children with autism and their peers. *Behavior analysis in practice*, 12(1), 78–94. <https://doi.org/10.1007/s40617-018-0228-x>
- Landa R. J. (2018). Efficacy of early interventions for infants and young children with, and at risk for, autism spectrum disorders. *International review of psychiatry (Abingdon, England)*, 30(1), 25–39. <https://doi.org/10.1080/09540261.2018.1432574>
- Lonigan, C. J., Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (2007). *Test of Preschool Early Literacy*. Austin, TX: PRO-ED, Inc.
- Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism Diagnostic Observation Schedule, Second Edition*. Torrance, CA: WPS.
- Mehling, M. H., & Tassé, M. J. (2016). Severity of autism spectrum disorders: Current conceptualization, and transition to DSM-5. *Journal of autism and developmental disorders*, 46(6), 2000–2016. <https://doi.org/10.1007/s10803-016-2731-7>
- Mullen, E. M. (1995). *Mullen Scales of Early Learning*. New York, NY: Pearson.
- Mundy, P., Delgado, C., Block, J., Venezia, M., Hogan, A., & Seibert, J. (2003). *Early Social Communication Scales (ESCS)*. Coral Gables, FL: University of Miami.
- Pavelko, S. L., Owens, R. E., Jr, Ireland, M., & Hahs-Vaughn, D. L. (2016). Use of language sample analysis by school-based SLPs: Results of a nationwide survey. *Language, speech, and hearing services in schools*, 47(3), 246–258. https://doi.org/10.1044/2016_LSHSS-15-0044

- Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). Classroom Assessment Scoring System (CLASS) Preschool Version. Baltimore, MD: Paul H. Brookes.
- Rowland, C. (2004). Communication Matrix. Retrieved from <https://communicationmatrix.org/>.
- Rowland, C., & Fried-Oken, M. (2010). Communication matrix: A clinical and research assessment tool targeting children with severe communication disorders. *Journal of Pediatric Rehabilitation Medicine: An Interdisciplinary Approach*, 3, 319–329. <https://doi.org/10.3233/PRM-2010-0144>
- Rutter, M., LeCouteur, A., & Lord, C. (2003). Autism Diagnostic Interview—Revised. Torrance, CA: WPS.
- Schopler, E., Van Bourgondien, M. E., Wellman, G. J., & Love, S.R. (2010). Childhood Autism Rating Scale, Second Edition. Torrance, CA: WPS.
- Sparrow, S., Cicchetti, D., & Saulnier, C.A. (2016). Vineland-III: Vineland Adaptive Behavior Scales, Third Edition. Minneapolis, MN: Pearson.
- Thiemann-Bourque, K. S., Brady, N., & Hoffman, L. (2019). Application of the Communication Complexity Scale in peer and adult assessment contexts for preschoolers with autism spectrum disorders. *American Journal of Speech-Language Pathology*, 28(1), 29–42. https://doi.org/10.1044/2018_ajslp-18-0054
- Wetherby, A., & Prizant, B. (2002). Communication and Symbolic Behavior Scales Developmental Profile. Baltimore, MD: Brookes.
- Yaruss, J. S., Max, M. S., Newman, R., & Campbell, J. H. (1998). Comparing real-time and transcript-based techniques for measuring stuttering. *Journal of Fluency Disorders*, 23(2), 137–151. [https://doi.org/10.1016/s0094-730x\(98\)00003-5](https://doi.org/10.1016/s0094-730x(98)00003-5)
- Yaruss, J. S. (1998). Real-time analysis of speech fluency. *American Journal of Speech-Language Pathology*, 7(2), 25–37. <https://doi.org/10.1044/1058-0360.0702.25>

Zimmerman, I. L., Steiner, V. G., & Pond, R. E., (2003). *Preschool Language Scale, Fourth Edition*.
San Antonio, TX: The Psychological Corporation.

Appendix A: Communication Complexity Scale

Communication Complexity Scale		
0	No response	
1	Alerting - a change in behavior, or stops doing a behavior	Preintentional
2	Single orientation only -- on an object, event or person; can be communicated through vision, body orientation, or other means.	Preintentional
3	Single orientation only + 1 other PCB (potentially communicative behavior); (e.g. staring at a toy and vocalizing)	Preintentional
4	Single orientation only + 2 or more PCBs (e.g. staring at a toy vocalizing and pointing to the toy)	Preintentional
5	Dual orientation -- shift in focus between a person and an object, between a person and an event using vision, body orientation, etc. (without PCB); (e.g. looking up from toy to examiner)	Preintentional
6	Triadic orientation-- eye gaze or touch from object to person and back; (e.g. looking from toy to examiner and back to toy within a short time frame)	Intentional Non-Symbolic
7	Dual orientation + 1 PCB (e.g. looking from toy to examiner and vocalizing)	Intentional Non-Symbolic
8	Dual orientation + 2 or more PCBs (e.g. looking from toy to examiner, vocalizing and pointing to toy)	Intentional Non-Symbolic
9	Triadic orientation + 1 PCB (e.g. looking from toy to examiner back to toy and vocalizing)	Intentional Non-Symbolic
10	Triadic orientation + 2 or more PCBs (e.g. looking from toy to examiner back to toy, vocalizing and waving at toy)	Intentional Non-Symbolic
11	One-word verbalization, sign or AAC symbol selection (e.g. selecting symbol for “ball” from a 4-choice array)	Intentional Symbolic
12	Multi-word verbalization, sign or AAC symbol selection (e.g. saying “more ball” to request the ball)	Intentional Symbolic
PCB stands for potentially communicative behaviors		

Appendix B: CCS Score Sheet for Play Assessment

Communication Complexity Scale: Score Sheet for Scripted Child Assessment

Date Scored: _____ Participant ID: _____ Your Name: _____ Other Coders: _____ Scored (Circle): INVIVO LIVE VIDEO

Task	Wind Up	Blocks	Snack	Music	Hammer	Fan	M-Tiles	Dots	Bubbles	Books	Bumble	Ball Toy
Start Time												
Stop Time												
Behavior Score 0-12												
Function BR JA RQ For scores of 6-12												
Mode SGD SP G Sign For scores of 11-12												

Score	Behavior
0	No Response
1	Alerting – A change in behavior, or stops doing a behavior
2	Single Orientation Only – A visual or body shift toward an object, communication partner, or activity
3	Single Orientation + 1 PCB (Potentially Communicative Behavior)
4	Single Orientation + 2 or more PCBs
5	Dual Orientation – Two-point look or body orientation, person-object/activity, object/activity-person, or person-person
6	Triadic Orientation – Three-point gaze shift, either between object-person-object or between person-object-person
7	Dual Orientation + 1 PCB
8	Dual orientation + 2 or more PCBs
9	Triadic orientation + 1 PCB
10	Triadic Orientation + 2 or more PCBs
11	One-word Verbalization, sign or AAC symbol selection
12	Multi-word Verbalization, sign or AAC symbol selection

PCB = Potentially communicative behavior such as gesture, vocalization or body movement.

Code	Function For scores of 6 through 12
BR	Behavior Regulation – Requesting or protesting an object, action, or activity
JA	Joint Attention – Used to direct the communication partner’s attention to an object, event, or activity
RQ	Response to Question – when the child is responding to a novel question

Code	Mode For scores 11 and 12
SGD	Speech Generating Device – Any high-tech AAC device with speech output
SP	Speech – Verbal speech sounds that form an intelligible word
G	Graphic Picture Symbol – PECS or visual schedule icons, pictures
Sign	Sign – Must fit outlined requirements for sign from manual, ASL, exact English signs

Notes: