

THE CONNECTICUT
DOCUMENTATION & OBSERVATION
FOR TEACHING SYSTEM (CT DOTS):

INITIAL RELIABILITY

June 2019

Background Information

The Connecticut Documentation and Observation for Teaching System (CT DOTS) is a framework to guide early care and education providers in a process of monitoring children's progress on the skills, abilities and behaviors in the Connecticut Early Learning and Development Standards (CT ELDS). CT DOTS was conceptualized and designed by the Connecticut Office of Early Childhood (OEC) in partnership with the University of Connecticut and an advisory group between 2016 and 2018. *The Connecticut Documentation and Observation for Teaching System: Instrument Development and Pilot Report* (Goldstein, 2017) outlines the development of CT DOTS, including the piloting of an initial version of this tool and recommendations for next steps. Recommendations from the pilot were used to refine CT DOTS and information about the expansion and field testing of the revised version are outlined in *The Connecticut Documentation and Observation for Teaching System: Field Test Report and Birth to Age One Expansion* (OEC, 2018). As a continuation of the overall examination of the psychometric properties of CT DOTS, an initial examination of rater reliability was performed in 2019. This report details that effort.

Initial Reliability

Planning

As a part of the ongoing examination of validity evidence for CT DOTS, it was determined that a study of the reliability of ratings would be particularly helpful in exploring the potential use of CT DOTS in Head Start programs and in conjunction with the Early Childhood Outcomes Reporting required for the Individuals with Disabilities Education Act, Section 619 services (preschool special education). In considering all of the available options for performing such as study, the following factors heavily influenced the final decision regarding the most rigorous approach under the current circumstances:

- As a new program, the total population size of CT DOTS trained practitioners was relatively small.
- Programs fully utilizing CT DOTS were geographically distributed across the state.
- There were limited opportunities for shared observations that would assure complete review of all age bands.

Given these constraints, the following plan was devised:

1. Construct a web portal supporting assembly of observational data sets and providing access to the observational data sets.
2. Author a series of observational data sets using text, video, or other available artifacts to serve as stimuli for scoring the level of child progress through a sequence of progressions. Data sets provide an opportunity for scoring across all domains and should be representative of infant, toddler, and preschooler age bands.
3. Develop an instrument [score sheet] to support the scoring of the observational data sets.
4. Recruit volunteers who had some training and experience with CT DOTS and/or with similar assessments to score some portion of the web-based observational data sets using the score sheet. Invite volunteers to participate, providing scoring directions, score sheets, and the method of submission.

5. Perform a cursory reliability analysis. Assure there were no contra-indications of inter-rater reliability after accounting for chance [Fliess' Kappa > 0].

Web Portal

During mid-fall in 2018 a web-based portal was constructed, providing a facility for the upload, storage, and review of observational data sets. Because this portal has utility in future trainings and additional reliability studies, the portal is described in detail here.

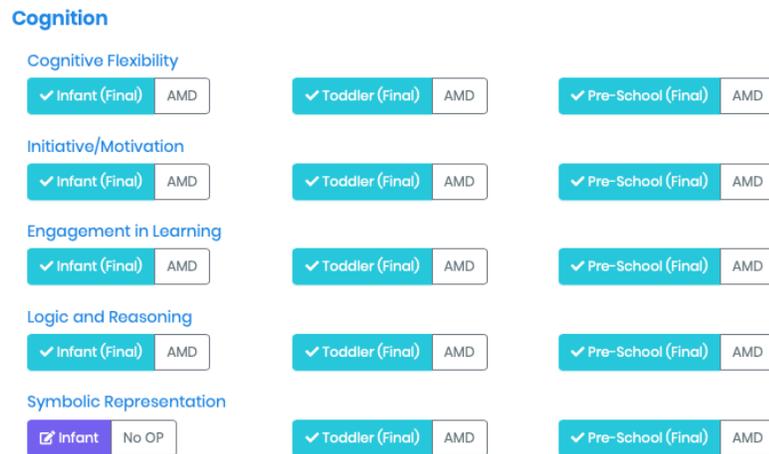
This system was designed as a LAMP application, written in PHP, delivered to the web with Apache with data storage in MySQL, all hosted on a Linux server. The access model supports three user roles: authors, readers, and administrators.

Table 1: Portal Access Privileges

Role	Privileges	Distribution
Author	Read access to the domain overview panel, read and write access to data set editing panel.	Single account
Reader	Read access to the domain overview panel, read access to formatted data sets.	Single account
Administrator	ALL	Single account

Primary navigation is done through a domain overview panel, a layout of domains displaying the progressions and age bands. Each button indicates the primary author of its associated data set. Certain progressions don't apply to each age band, indicated by a color change and the text "No OP," as shown in Figure 1., as is the case of symbolic representation in the Infant age band. The button also indicates the stage of editing for each data set, one of either "Initial," "Review," or "Final."

Figure 1: Domain Overview Panel



Authors were able to populate data sets in the data set editing panel (see section on Data Set Construction). Authors could select the applicable age band for the progression, which dynamically retrieved the associated descriptive text, as seen in Figure 2.

Figure 2: Data set Age Band Selection

Cognition - Cognitive Flexibility - Infant

Skill Description and Examples

Age Band 1	Age Band 2
6-9 Months	9-12 Months
<p>Uses varied actions to explore and interact with their environment</p> <p>Example: Seeks object that is partially covered (e.g. moves blanket to retrieve object) Responds differently to adults and children Acts to have enjoyable activity repeated (e.g. coos to get adult to bounce him/her again, touches toy to get it to repeat noise)</p>	<p>Adjusts actions to accomplish a desired effect</p> <p>Example: Uses different actions to obtain objects (e.g. jumps to reach something) Adjusts actions to maintain objects in a parti</p>

The next step in data set construction is providing descriptive text. In the case shown in Figure 3. text is provided describing evidence from planned experiences, naturalistic observation, and family input.

Figure 3: Data Set Text Input

Data Set

Evidence Text

Rich text editor toolbar: Bold, Italic, Underline, Text Color, Background Color, Bulleted List, Numbered List, Indent, Outdent, Undo, Redo, Styles, Normal, etc.

Evidence 1 (planned experience)
Video below, please watch from :30 to 1:40 of the video.
Jacob explored toy in multiple ways.

Evidence 2 (naturalistic observation)
While eating his snack Jacob was trying to put his cup in his bowl. The first time he tried to do this he tipped his bowl over. He straightened his bowl and tried again. This time I showed him how to hold the bowl. He watched me but did not imitate holding the bowl.

Evidence 3 (family input)
Response on Family Input Booklet for 6-9 month age band:
How do you know when your child is interested in something?
He looks at it, reaches for it, calls out to us or his sister if she has it.
How does he let you know he wants something to continue?
When I stop doing something he likes he looks at me, sometimes he starts crying then I know he wants more.

body p

Links to non-textual artifacts can also be provided. In this case shown in Figure 4., a link is provided to a video segment.

Figure 4: Data Set Artifact Linking

Link 1
<https://youtu.be/235CwZLFerU>

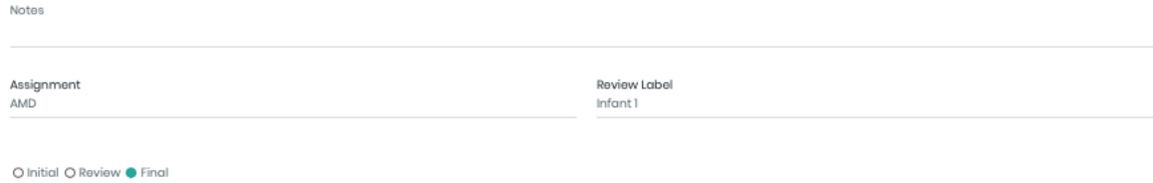


Link 2

Link 3

As shown in Figure 5., The final data fields support notes, the primary author assignment, and the editing stage.

Figure 5. Additional Data Set Fields



Once data entry was completed for all data sets the elements can be presented to the readers in a unified noneditable view.

Data Set Construction

Efforts were made to gather authentic observation descriptions to include in the data sets for review; however, given the limited implementation of CT DOTS at the time of this study it was not feasible to populate the portal with actual examples. Therefore, early childhood experts with a high level of familiarity with CT DOTS were recruited to construct the scenario data sets for inclusion in the online portal.

Table 2: CT DOTS Data Development

Data Set Developer	Experience with CT DOTS	Education
1	CT DOTS trainer and coach	<ul style="list-style-type: none"> • M.S. Infant and Parent Development, N-6 cert. • MBA • Early Childhood Leadership Certificate • B.A. Dance
2	CT DOTS development and implementation	<ul style="list-style-type: none"> • Ph.D. Educational Psychology • M.Ed. Early Childhood Special Education • B.S. Child Psychology
3	CT DOTS trainer and coach	<ul style="list-style-type: none"> • M.S. Special Education • B.A. Psychology

Experts each took primary responsibility to author data sets for a number of CT DOTS Observation Progressions. Each data set consisted of three pieces of data, including videos and/or photographs and descriptions of child behavior/skills. Videos and photos were drawn from a variety of publically available sources. Each data set was aligned to a particular age band, with an adjacent age band specified for inclusion as a possible response. Each data set and assigned age bands were subsequently reviewed by another member of the team who either edited or approved the data set.

Score Sheet

In order to reduce the complexity of the online activity, it was decided that the scoring for each data set would be completed via a paper score sheet to be scanned and emailed or mailed upon completion.

Score sheets were created that required a simple check to indicate the level of child progress within the two age bands specified. For each of the two age bands presented there were two possible responses, emerging skills or consistent skills. This study design meant that there were 4 possible responses for any given data set.

The following rubric was provided to guide participant responses.

Emerging (E)	The described behaviors have been observed in some situations or settings and/or occur with adult support but are not used consistently.
Consistent (C)	The described behaviors are exhibited consistently across multiple settings with limited adult support.

A score sheet excerpt can be seen in Figure 6. Additionally, the sheet included three demographic questions concerning role, experience, and education. The score sheet, demographic questions, and directions for participation were combined in a single PDF document for each age band, infant, toddler, and preschooler. Complete Instruments can be found in Appendix A.

Figure 6: Score Sheet Excerpt

CT DOTS Reliability Data Collection Form											
Infant											
Domain	Skill Area	Data Set	0-3 months		3-6 months		6-9 months		9-12 months		KEY E=Emerging C=Consistent
			E	C	E	C	E	C	E	C	Comments
Cognition	Cognitive Flexibility	INFANT 1									
	Initiative/Motivation	INFANT 2									
	Engagement in Learning	INFANT 3									
	Logic and Reasoning	INFANT 4									

Participation Recruitment and Data Collection

The only criteria for participation as a scorer was some degree of experience with the CT DOTS framework. Invitations were sent via email to state-funded center contacts, CT DOTS trainers, and conference attendee rosters. Given that most practitioners would not have experience across the entire spectrum of age bands participants were allowed to choose the age bands they were most comfortable with, infant, toddler, and/or preschooler.

By the end of the scoring period in winter 2019, 25 score sheets had submitted. Of the 25, two were removed from the set for diverging from the completion instructions. The final counts were 2 infant sheets, 3 toddler sheets, and 18 preschooler sheets.

Tables 3-5 below illustrate the demographic distribution of participants. Note that row totals are not valid, as participants may have submitted score sheets for more than one age band.

Table 3. Participant Demographic Distribution: Role

ROLE/AGE BAND	Infant	Toddler	Preschooler
Teacher	-	1	8
Assistant Teacher			
Program Administrator	2	2	6
Consultant or coach	-	-	1
Other	-	-	3
TOTAL	2	3	18

Table 4. Participant Demographic Distribution: Experience

EXPERIENCE\AGE BAND	Infant	Toddler	Preschooler
<5 years	-	-	1
5-10 years	-	-	1
10-15 Years	-	-	3
15+ years	2	3	13
TOTAL	2	3	18

Table 5. Participant Demographic Distribution: Education

EDUCATION\AGE BAND	Infant	Toddler	Preschooler
Child Development Associate (CDA) Credential	-	-	-
Associate's Degree	-	-	-
Bachelor's Degree			3
Master's Degree	2	3	13
Doctoral Degree	-	-	2
TOTAL	2	3	18

All responses were coded and compiled into three distinct data sets by age band.

Inter-Rater Reliability

The infant age band had two (2) raters and 16 cases, where the cases represent the number of infant appropriate progressions. The percentage agreement at a tolerance level of 1 was 87.7%. Fliess' Kappa

was selected as the measure of reliability, understanding that it does not reduce to Cohen’s Kappa, and does estimate slightly higher. Fleiss’ Kappa was 0.138 ($p = 0.412$). Complete demographic homogeneity of the scorers and small n for this age band prevented any group analysis.

The toddler age band had three (3) raters and 24 cases, where the cases represent the number of toddler appropriate progressions. The percentage agreement at a tolerance level of 1 was 50%. Fleiss’ Kappa was 0.204 ($p = 0.00352$). Substantial demographic homogeneity and small n for this age band prevented any group analysis.

The preschool age band had eighteen (18) raters and 25 cases, where the cases represent the number of toddler appropriate progressions. The percentage agreement at a tolerance level of 1 was 8%. Fleiss’ Kappa was 0.158 ($p = 0.00$). The number of scorers and demographic heterogeneity allowed for a degree of group analysis, shown in Table 6. Small n and unsuitable distributions prevented examination of interactions.

Table 6. Preschooler Age Band Intra-Subgroup Reliability

Subgroup	n	Fleiss’ Kappa (p)
15+ years of experience	13	0.162 (0.00)
MS degree or higher	15	0.145 (0.00)
BS degree	3	0.050 (0.46)
Ph.D.	2	0.200 (0.08)
Teacher	8	0.098 (0.00)
Administrators	6	0.15 (0.001)

Summary and Recommendations

Under the conditions of this current study reliability of agreement scores (Fleiss’ Kappa) are weak but indicate reliability better than chance, given that p was greater than zero for all subgroups. Several factors likely influenced these initial reliability results. For each of the likely influences below, recommendations have been provided for future studies in order to gather more evidence regarding how great reliability across ratings can be achieved using CT DOTS.

- The nature of the data sets used as rater stimuli: For future CT DOTS reliability investigations, using data collected in classrooms over time would provide realistic stimuli which would better match the kinds of information use when making rating determinations.
- The conditions under which ratings determinations are made: The conditions in this initial reliability study (determining child performance given a fixed data set) did not provide all affordances available in a direct-observation setting. When using CT DOTS, teachers are able to accrue additional natural observations, planned experience, or family Input in situations where the child’s performance levels cannot be confidently determined given immediately available data. In addition to the formal data collection process, teachers using CT DOTS typically have many informal observations that help them to understand if the data samples are typical of a child’s skills or behaviors.
- The amount of training on CT DOTS: This study did not involve any specific training designed to promote consistency of ratings between providers using CT DOTS. Future studies should include

training on the use of CT DOTS and include ample opportunities for reviewing data, discussing ratings, and working toward inter-rater reliability. Where reportable, intra-subgroup reliability scores in the current study indicated that reliability was higher with increased education and experience, which is consistent with observation in practice. Further exploration of the relationship of targeted training, education, experience and the reliability of ratings is warranted as a part of all future studies.

- The dimensions represented in CT DOTS: CT DOTS was developed based on existing domains and progressions within the CT Early Learning and Development Standards. A future factor analysis would provide additional evidence related to validity and would allow for an examination of progression uni-dimensionality for reliability.

This study only preliminarily examined overall inter-rater reliability. Intra-rater reliability or other forms were not considered. The model of inter-rater reliability employed assumed all individual items were part of the same scale, which was not confirmable within this study. It could not consider reliability by subscale.

Appendix A

Score sheets