

Measuring the Implementation of UDL in Classrooms and Schools: Initial Field Test Results

Remedial and Special Education
2020, Vol. 41(4) 231–243
© Hammill Institute on Disabilities 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0741932520908015
rase.sagepub.com



James D. Basham, PhD^{1,2}, J. Emmett Gardner, PhD³,
and Sean J. Smith, PhD¹

Abstract

The design, flexibility, and iterative nature that is inherent to Universal Design for Learning (UDL) makes it difficult to consistently measure. With federal policy encouraging the implementation of UDL, there is an increased need for practitioners to reliably measure the occurrence of UDL. The UDL Observation Measurement Tool (UDL-OMT) was developed to measure UDL implementation in classrooms. This article presents the conceptual underpinnings of UDL measurement and the results of an initial field test. Results indicate that the UDL-OMT has good to excellent internal consistency and can characterize differences in UDL implementation across a continuum of settings. Discussion focuses on the reliability of the UDL-OMT and its potential as a formative evaluation tool for practitioners and school-based personnel. Additional considerations include promising research applications and how the nature and context of classroom instructional factors as well as observers' UDL knowledge influence interpretations of observations.

Keywords

Universal Design for Learning, instruction, classroom, evaluation, instructional evaluation

Introduction

As education leaders, curriculum vendors, technology developers, researchers, and education change agents continually introduce the next best concept, widget, or practice, educators on the ground require a broad framework to support effective decisions in designing across the needs of all learners. UDL has been researched, advocated for, and embedded in education policy (ESSA, 2015), and guidance (National Educational Technology Plan [NETP], 2017), as a framework for proactively designing learning environments for all students (Basham & Marino, 2013; Meyer et al., 2014). The framework contains three principles, nine guidelines, and 31 “checkpoints” representing design considerations (formerly, Center for Applied Special Technology [CAST], 2018; see Figure 1). However, as UDL is a framework for designing learning environments and experiences rather than a specific practice, how to reliably measure the implementation of UDL has challenged researchers and practitioners for some time (see Basham et al., 2016; Basham & Gardner, 2010; Edyburn, 2010; Rao et al., 2014). The challenge to reliably measure UDL's implementation continues to present issues concerning the operationalization of its practice.

As a framework, UDL integrates and classifies variables associated with the design of effective learning environments

for all learners, including students with disabilities. Generally, UDL is focused on ensuring learning environments and associated experiences are designed to support access, build understanding of knowledge and skills, and help learners internalize behaviors that support expert learning (Meyer et al., 2014). The UDL framework supports the variability inherent in all learners by promoting proactive design of learning experiences and overcoming unneeded barriers to learning. The focus is to provide options across multiple means of engagement, content representation, and actions that demonstrate understanding.

Within the United States, UDL is identified in the Every Student Succeeds Act (ESSA, 2015), as a foundational framework to support educating all learners. Although UDL has been around roughly 20 years, it is one of the least understood concepts in education and special education (Edyburn, 2010; Shah, 2012; Zhang et al., 2020). Nearly

¹University of Kansas, Lawrence, USA

²CAST, Wakefield, MA, USA

³The University of Oklahoma, Norman, USA

Corresponding Author:

James D. Basham, Department of Special Education, University of Kansas, Joseph R. Pearson Hall, 1122 W Campus Rd., Lawrence, KS 66045, USA.

Email: jbasham@ku.edu

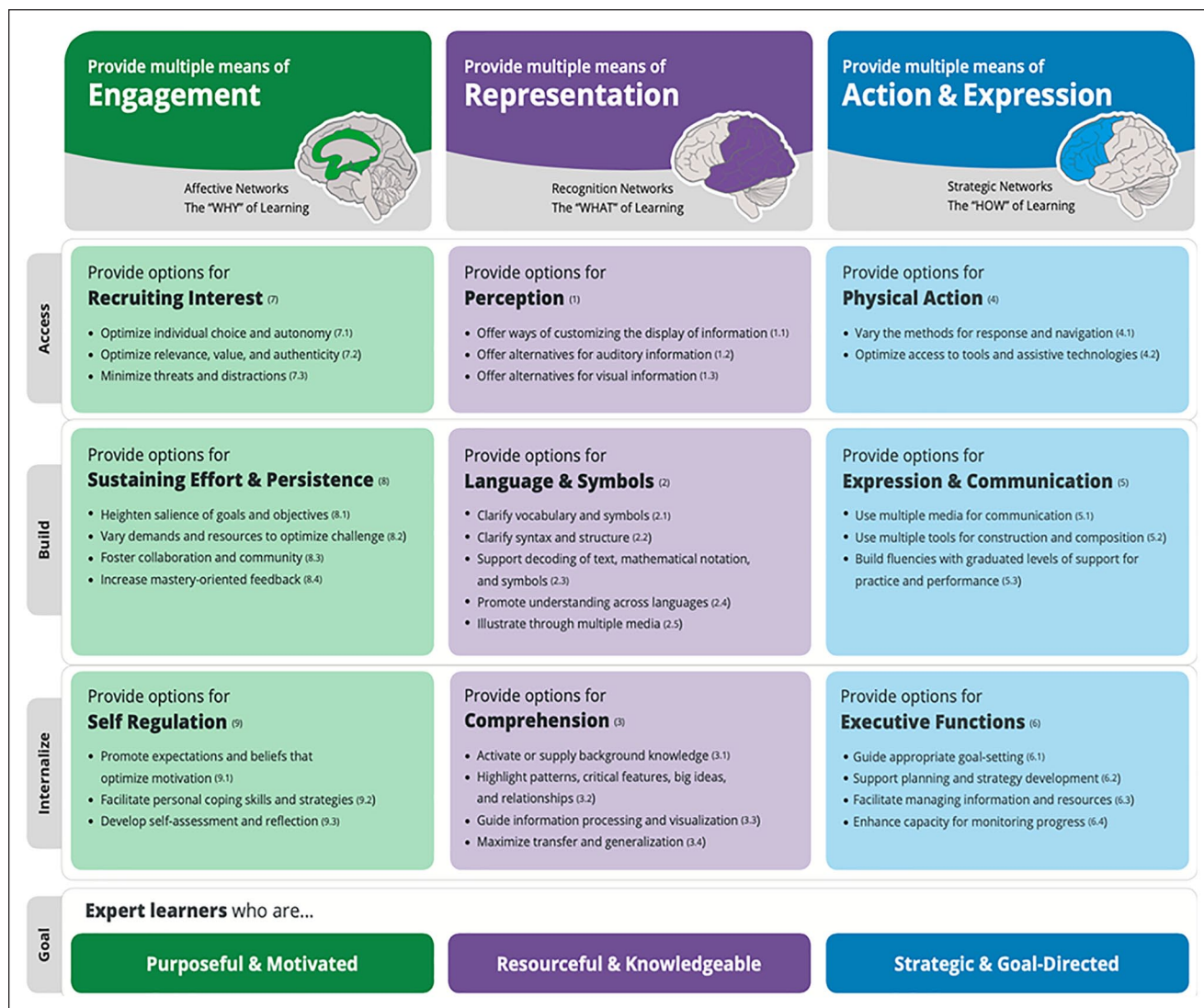


Figure 1. Universal Design for Learning (UDL) Guidelines (CAST, 2018).

Source. Used with permission from CAST (2018).

every state is attempting to implement some form of UDL (National Center on UDL, 2012).

For those of us who are often called upon to provide support for state- and district-level implementation, a common misperception is that UDL is a simple practice or even a strategy that can be quickly implemented to positively affect student outcomes. Another misperception is that UDL is simply related to the use of instructional technology or that UDL is just for students with disabilities (Basham & Gardner, 2010; King-Sears, 2009). Educators desiring to implement UDL with some level of fidelity and clear outcome effects also vary in their definition of UDL and how it should be measured (Basham & Gardner, 2010; Edyburn, 2010; Rao et al., 2014).

Rose and Meyer (2000, 2002) discussed UDL as a new way to look at and approach curriculum planning, instructional practice, and assessment. As highlighted by Edyburn (2010), many professionals who understand UDL typically think of it in the context of its three primary principles:

- Multiple Means of Engagement, where instruction embeds a range of strategies to engage the interests and motivations of all students;
- Multiple Means of Representation, where information is presented and represented in ways that account for all sensory and learner variability of students;
- Multiple Means of Action and Expression, where students have a choice of what medium she or he will

use to physically access content, technology (as a tool), and communicate knowledge acquisition.

An expanded understanding of UDL, the principles, and the interworking of the framework itself can be found in Meyer et al. (2014) and through use of the UDL Guidelines (CAST, 2018).

In reality, the underlying principles, associated guidelines, and checkpoints of UDL (see Figure 1) provide a highly researched and intricate framework for supporting variability associated with learning in an environment. Across design considerations, the framework is focused on supporting access, building knowledge and skills, and encouraging internalization of knowledge and skills in relation to becoming an expert learner (CAST, 2018; Meyer et al., 2014). Beyond these general aims, each of the principles, guidelines, and checkpoints are supported by research in multiple disciplines and fields (Meyer et al., 2014; Tokuhamma-Espinosa, 2010). For example, Basham et al. (2018) provide a summary of the research aligned with the UDL Guidelines (CAST, 2018) across mind, brain, and education research. The meta-analyses of Hattie (2011, 2012) attribute significant effect sizes to a variety of variables that are synonymous with many of the UDL checkpoints. In addition, research on UDL has begun to show positive outcomes across multiple studies (e.g., Basham et al., 2010, 2011; Kennedy et al., 2014; Marino, 2009; Proctor et al., 2011; Rappolt-Schlichtmann et al., 2013).

In a systematic review of the literature, Rao et al. (2014) found that the UDL framework was beginning to demonstrate positive outcomes across literacy, math, and science content as well as positive effects on overall instructional engagement. However, although positive effects continue to emerge in research, there is still a lack of a consistent operationalized understanding of what constitutes UDL implementation (Ok et al., 2017), particularly for practitioners and other school-based personnel. As highlighted in these recent literature reviews, studies vary on how UDL is operationalized in practice (Ok et al., 2017; Rao et al., 2014).

The lack of a consistent means for measuring UDL within a specific practice or some research interventions is concerning (Basham & Gardner, 2010, 2016; Edyburn, 2010; Ok et al., 2017; Smith et al., 2019). Suffice it to say that both researchers and practitioners operationalize UDL in a variety of ways. It is problematic that a practice somewhat aligned with any of the three principles, or misconception of the framework, can be identified as UDL. The lack of a stable understanding of UDL leaves those attempting to implement it confused about what constitutes UDL and how to apply and measure it. The purpose of this article is to describe the development and initial validation of an observation tool, to be used by practitioners and school-based personnel, for measuring UDL implementation.

Obstacles in Measuring the Implementation of UDL

The focus on flexibility, proactive, and iterative design inherent to UDL makes it difficult to consistently measure. When measuring, it is also necessary to differentiate UDL-aligned practice from other practices, such as effectively integrating technology and differentiated instruction (Basham & Gardner, 2010; Hall et al., 2003; King-Sears, 2009). Planning for all learners through the integration of appropriate goal setting, deciding on and using various instructional strategies and materials, as well as supporting multiple means of assessment requires more than changing a few aspects of the learning environment.

As discussed by Basham and Gardner (2010), the complexity of measurement was apparent during implementation of a UDL environment that transformed an ineffective urban school into an effective UDL-based science, technology, engineering, and mathematics (STEM) school (see Basham et al., 2011). Confusion occurred when outside evaluators were attempting to attribute the school's positive transformation to UDL. Specific questions emerged about how to identify whether UDL was the driving force in the school's change or whether another set of variables led to the change. At that time, the evaluators, who had limited understanding of UDL, characterized their impression of UDL as a Gestalt effect (e.g., when the teachers had an "ah-ha" moment, where they perceived the sum of individual changes due to use of UDL checkpoints to have a miraculous effect on improvement of STEM-related knowledge, skills, and attitudes). Basham et al. (2011) suggested that the implementation of the UDL-based STEM activities may have provided a Gestalt effect to the observers, but in reality, a good deal of purposeful planning occurred as well as use of multiple strategies, which could be replicated at other schools. Operationalizing this effect was challenging because, although outcomes were clear, UDL's implementation was difficult to measure.

Conceptually, UDL has strong face validity. The individual/discrete principles, guidelines, and checkpoints of UDL have been validated with regard to experimental and quantitative evidence as well as scholarly reviews and expert evidence (see <http://udlguidelines.cast.org/more/research-evidence>). However, when it comes to operationalizing and measuring UDL as a single complete construct, the enormity of UDL elements complicates matters (Basham & Gardner, 2010). The current UDL Guidelines (CAST, 2018) contain three principles, nine guidelines, and 31 checkpoints that represent design considerations (and discrete elements) that constitute the framework. Herein lies one of the principal dilemmas for operationally defining and measuring UDL. Does UDL require that a minimum number and/or combination of specific guidelines and/or checkpoints be present? Is this combination of guidelines

and/or checkpoints static or fluid? UDL guidelines and checkpoints are not representative of a checklist (Basham & Marino, 2013). UDL is about a goal-driven design that requires addressing learner variability. The UDL Guidelines (CAST, 2018) provide an understanding of the underlying variability that should be considered in the design of a learning environment or experience. At the same time, according to Edyburn (2010), UDL is not a naturally occurring phenomenon—there needs to be clear evidence of intentional a priori consideration and classroom application of instructional flexibility, strategies, and technology (defined as tools) to enhance learning.

Another complicating factor in the measurement of UDL, especially in active learning environments, is that design is iterative and continually changing, based on context (e.g., goals of instruction, time available, learner needs, and tools available). Similar to other types of design research in education (e.g., Cobb et al., 2003), UDL has a direct association with the design of learning environments (Basham et al., 2011; Basham & Marino, 2013). As a result, every time one of these learning environments is observed, only a snapshot is captured across an iterative time series that should be working toward continuous improvement (Bryk et al., 2011, 2015; Lewis, 2015). Thus, during a practice-based implementation of UDL (e.g., Basham et al., 2011), as contrasted to a priori experimental UDL-based interventions (e.g., Kennedy et al., 2014; Marino et al., 2014), observation-based tools attempting to capture UDL in everyday classroom environments must account for the possibility of a more dynamic implementation of UDL. Therefore, any methods, tools, and associated products (e.g., publications and presentations) should consider the relationship with design research and not as a single-shot summative assessment.

Overall, there is a need to measure the implementation of UDL in learning environments. Understanding this need, a project was undertaken to develop and validate an instrument for observing UDL-based environments. This project was driven to answer two questions:

Research Question 1: Can an observation tool be developed and validated that reliably measures UDL implementation within a classroom setting?

Research Question 2: Can a UDL observation tool distinguish implementation differences across instructional environments?

Method

Description of UDL Observation Measurement Tool (UDL-OMT)

The UDL-OMT is a 42-item assessment tool designed to measure the level of UDL alignment within an instructional

environment or experience. Developed over a 4-year period by the authors of this study and experts in UDL, in consultation with other experts in the field, including leadership at CAST and the founders of UDL, the UDL-OMT was constructed to provide researchers and evaluators a means to observe UDL in practice. The UDL-OMT can be used as a tool for observing a whole-class UDL implementation or a situation with a targeted group of students. It can be used across various instructional environments and learning experiences within a context of supporting continual iterative improvement.

The UDL-OMT is designed for observers familiar with the UDL framework and the knowledge that specific tools or strategies systematically occur in the context of supporting access, building knowledge and skills, and supporting internalization of understanding and skills across the implementation of the framework. The items in the UDL-OMT are aligned to the UDL Guidelines 2.0 (CAST, 2018), but are framed and written using language familiar to educators. See Table 1 for the relationship among the UDL Guidelines and items on the UDL-OMT.

The UDL-OMT is conceptualized as a semi-structured dynamic observation tool. Rather than organizing the observation around discrete principles, guidelines, or checkpoints, the tool was designed based on the general flow of a lesson, identifying places where particular checkpoints would more likely be observed based on their relevance to instructional events. Although the UDL-OMT considers the design of the learning environment relative to teacher implementation, it also recognizes that students' use of strategies and tools, as well as how students respond to the environment, must be flexible. Observers adjust their rating of UDL based on their ongoing data collection during the observation. Individual items are scored using a scale of 0 (*no evidence of UDL*) to 3 (*dynamic, interactive UDL*). Table 2 presents the detailed characterization of UDL relative to the 0 to 3 scale.

The UDL-OMT is used to consider alignment to UDL within the context of four sections: (a) introducing and framing new material (six items), (b) content representation and delivery (nine items), (c) expression of understanding (seven items), and (d) activity and student engagement (nine items). In addition, each section provides observers opportunities to comment on their perceptions of the effectiveness of implementation. Finally, the tool concludes with a group of items that record observers' overall impression of the variety of UDL principles that were perceived to be effectively implemented during the observation and a set of items asking observers to reflect on students' overall engagement, interest, and task focus during the lesson. Based on observed activities, observers are free to choose what sections of the tool to use. For instance, if observers witness the second day of a 5-day unit, they may not observe the framing of new content; thus, they would not use that section. This article reports on the reliability of three sections that consider the

Table 1. Corresponding UDL-OMT Sections and Items Mapped to CAST (2018) UDL Guidelines and Checkpoints.

Content representation and delivery	
<ul style="list-style-type: none"> • Presentation of information allows for customization (G1; C1.1). • Instruction allows alternatives for visual display of information (G1; C1.3). • Instruction allows alternatives for auditory information (G1; C1.2). • Supports options for multiple languages (G2; C2.4). • Supports multiple levels of content understanding (e.g., novice, intermediate, or expert) (G3; C3.1, C3.3). • Supports understanding of relationships across disciplines, settings, or concepts (G3; C3.2). • Clarifies content-specific vocabulary, symbols, and jargon (G2; C2.1, C2.3). • Clarifies content-based syntax and structure (G2; C2.2). • Highlights options for self-directed clarification of vocabulary and symbols (G2; C2.1, C2.3). 	
Expression of understanding	
<ul style="list-style-type: none"> • Allows options for learners to express understandings in a variety of ways (G5; C5.1, C5.2, C5.3). • Provides access to a variety of tools and/or technologies for students to express their understanding (G4; C4.1, C4.2). • Builds competencies in the use of multiple options for expressing their understanding (G5; C5.3). • Intentionally provides supports for students' problem-solving and critical-thinking abilities (G3; C3.2). • Provides options that guide students to plan, develop strategies, and/or goal-setting that promotes expression of understanding (G6; C6.1, C6.2). • The environment facilitates management of information and resources to achieve desired learning outcomes (G6; C6.3). • Facilitates student self-monitoring of progress (G6; C6.4). 	
Activity and student engagement	
<ul style="list-style-type: none"> • Promotes learner choice and self-determination while engaging with the content (G9; G7; C7.1). • Provides a variety of activities relevant to all learners (G7; C7.2). • Promotes sustained effort and focus (G8; C7.3). • Encourages learners' use of strategic planning to complete instructional tasks (G6; 9; C6.2, 9.2). • Encourages collaboration and communication among learners (G8; C8.3). • Supports multiple levels of challenge (G8; C8.2). • Provides for self-reflection and self-assessment (G9; C9.3). • Provides formative progress monitoring and content checks (G6; 9; C6.3, C6.4, C9.3). • Provides closure that reiterates big ideas and instructional purposes (G8; C8.1, C8.4). 	

Note. UDL Guideline (G), UDL Checkpoint (C). The items map to the UDL Guidelines (CAST, 2018) presented in Figure 1. UDL = Universal Design for Learning; OMT = Observation Measurement Tool.

Table 2. UDL-OMT Rating Scale Description and Examples.

Rating	Description and examples
0 = <i>no evidence of UDL</i>	No multiple means—Learners were not provided options for engagement, expression, or learning. Learners experienced or used a single strategy or tool; they did not experience or engage in any other option to access the information, demonstrate understanding, take action, or engage in the activity.
1 = <i>incomplete evidence of UDL in environment</i>	Learners experienced or used a single strategy or tool; however, other strategies or tools were clearly available in the environment but not explicitly used by students or referenced by the teacher.
2 = <i>UDL is occurring</i>	Learners experienced or used at least two or more strategies and/or tools for how they accessed the information, demonstrated understanding, took action, or engaged in the activity. However, the options were generally static and/or traditional in nature; learners are relying heavily on teacher dependence to support the interpreted range of learner variability.
3 = <i>dynamic, interactive UDL</i>	Learners experienced or used at least two or more strategies and/or tools for how they accessed the information, demonstrated understanding, took action, or engaged in the activity; in addition, learners are relying far less on teacher dependence, have more choice on their use of tools or strategies that are customizable, interactive, dynamic, and efficiently support the interpreted range of learner variability.

Note. UDL = Universal Design for Learning; OMT = Observation Measurement Tool.

alignment of UDL related to content representation and delivery, expression of understanding, and activity and student engagement ([b] through [d]).

Development of the UDL-OMT

The development of the UDL-OMT followed an iterative design process organized across four design phases. Phase 1 focused on the initial instrument design. During this phase, the first and second authors, with feedback from other researchers and educators who possess a strong understanding of UDL, discussed and worked to a consensus on the identification of UDL terms, item wording/verbiage, and item scaling. The UDL-OMT was designed for use by individuals with familiarity concerning the UDL Guidelines (CAST, 2018) and the tenets of UDL implementation (Basham & Marino, 2013). Particular attention was directed at identifying language that would accurately translate the more psychoeducational/technical wording of the principles and checkpoint of UDL into phrases that would be more understandable by teachers. For example, the UDL Checkpoint 2.2 “Clarify syntax and structure” (CAST, 2011, 2018) was eventually translated to “Clarifies content-specific vocabulary, symbols, and jargon.”

Once a beta version of the instrument was completed, Phase 2 of the development consisted of testing the feasibility and usability of the UDL-OMT in face-to-face situations. During this phase, the designers used the instrument in two elementary classrooms, one middle school classroom, and one high school classroom. Specifically, this test was focused on whether the tool could be reasonably used in real-world situations and where issues of usability could be identified.

Phase 3 of the design focused on modifying the instrument after the initial usability test. Primary findings from Phase 2 of the usability test indicated that the instrument was too complex to use in a real-world classroom environment. Thus, the first and second authors worked to modify the flow of the instrument to make it easier to use in classroom settings. During Phase 3, the validated Reformed Teaching Observation Protocol (RTOP; Piburn & Sawada, 2000) was identified and used as a reference model for the flow of the instrument. During Phase 3, two other procedural and implementation issues were made by the first and third authors based on observing videos of UDL lessons recorded in K–12 classrooms. These adjustments included providing clarity to wording, deleting redundant items, and establishing observation procedures based on the new flow of the tool. Toward the end of Phase 3, the researchers formally and informally shared the tool with colleagues in the UDL community. Throughout the multiyear development process, the authors sought feedback from practitioners and researchers (>10) who both published and presented on UDL, including leadership at CAST and K–12 teachers who were working in districts known to be implementing UDL.

Phase 4 of the design process consisted of conducting observations in 11 UDL-aligned classrooms for an initial evaluation or field test of the tool. Phase 4 is the focus of this article.

Observation Sites

With the focus on whether UDL was taking place and how it might be improved, the evaluators conducted observations in 11 classrooms for a total of 22 observations. These classrooms are located in a school district that is known across the UDL community for supporting implementation. For instance, the school district has UDL embedded within educator evaluations and ensures that students are supported in UDL environments prior to referring for special education services. The district is located in a midsize city in the Midwest of the United States. According to Wolfram Alpha (2018), the surrounding community has a population of 77,870 residents, and the school district serves roughly 11,500 students. The school district’s student population comprises White (71.7%), Hispanic (15.2%), Asian (6.3%), African American (2.1%), Native American (0.3%), Native Hawaiian or Pacific Islander (0.1%), and multiracial (4.2%). Approximately, 45% of the students receive free or reduced-price lunch. Roughly, 13% of the students receive special education services.

Prior to observations, school district leadership identified classrooms depicting a continuum of UDL implementation, from little or no implementation to apparent implementation. Presumptions made by the school district leadership were not revealed to the observers or teachers. Teachers in the district had no prior knowledge about observations, other than there was general knowledge in the district that, over the course of 2 days, some classrooms would be visited. Observations occurred in 11 classrooms, ranging from Kindergarten to 12th grade. Four were at the Elementary school level (Grades K, 3, 4, and 5), four at the Middle school level (science, math, English, and self-contained classroom for students with behavior disorders), and three at the High school level (English, Spanish, and Literature). Each classroom was observed one time, with both observers present. The average time of individual observations was 40 min (for a total of 440 min of observation time), with a range of 20 to 55 min per classroom. The average time spent in a single classroom (which included a short meet and greet, as well as a short postobservation chat with the classroom teacher) across all visits was 45 min, with a range from 20 to 60 min.

Observations were conducted in seven schools. Two schools were considered specialized or separate from what is normally identified as an everyday school setting. The first of these settings was a school implementing a nationally recognized project-based learning model that integrated STEM. The second school was a 24-hr segregated living

and learning environment for individuals with behavioral/psychological disorders. Five schools were in what would be considered a neighborhood or city settings, and the other two schools were in rural settings.

Observer Roles

Two individuals, both with extensive knowledge of UDL (e.g., taught multiple teacher preparation and graduate classes in UDL; presented and published on UDL) and involved in the development of the UDL-OMT, were the observers. For the validation process, the two observers took on different roles throughout the observations (Schutt, 2018). One observer took on the view of being more open and accepting, presuming that UDL was present. This individual focused on the positive practices of UDL across the various instructional environments, with an eye toward UDL naturally taking place. The second observer approached observations and experiences from a more conservative stance. The second individual was intensely focused and critical of instructional practices, with an eye toward thinking UDL was likely to be observed only when there were clear and obvious events/activities that involved behavior and products indicative of a UDL checkpoint. Together, it was thought that the two observers, in purposeful roles (and being internally consistent in their observations), would provide a more realistic and balanced perspective toward whether UDL could be observed in the environment. All classrooms were observed by the two observers at the same time.

Observation Procedures

The process for accessing classrooms was consistent across all sites. Upon arriving at the school, the district liaison introduced the observer to the school's principal or building leader. The building leader asked the observers about their intent and needs for the observation, the impending observation time line, and then provided general information about the school and its unique mission (sometimes while walking around with the observers). The building leader typically walked the observers to a classroom and introduced them to the classroom teacher. The observers entered the room, found a place to sit, and began the observation. In all cases, teachers welcomed the evaluators and demonstrated obvious openness to being observed.

During an observation, each observer positioned themselves in different corners of the classrooms. At the beginning of each observation, approximately 5 min was used to observe the overall climate of the classroom, with no focus on scoring items on the instrument. After 5 min, each observer acknowledged to one another that the observation was to begin. The observers began rating the occurrence of UDL, shifting among Sections B through D of the instrument as needed. As classroom activities progressed and circumstances allowed, the

observers moved about the classroom, noting clusters of students working on group activities.

The observers focused on identifying the occurrence of UDL (as represented by the UDL-OMT items) via observation of student and teacher behaviors, use of instructional strategies, permanent products, and classroom tools that were suggestive or explicit of UDL. For example, they observed lectures or conversations between a teacher and student(s) where the teacher provided multiple examples to illustrate or clarify a concept or procedure. They took note of the nature and degree of interaction and dialogue among students working in small groups. They accounted for the presence (or absence) of relevant products (e.g., workbooks, worksheets, manipulatives, technology, and access to the internet) that were available or in use by students when engaged in learning activities. The amount of UDL for each item on the UDL-OMT was recorded based on the UDL-OMT rating scale (see Table 2).

Both observers completed the UDL-OMT independent of each other, never coming in close proximity to one another or sharing information. During observations, observers also compiled an anecdotal record to document specific and/or unique examples of UDL, or to highlight situations where they thought UDL could easily have been applied but was omitted.

Scoring Procedures

Throughout observations, the observers accessed either a paper or digital version of the instrument. Observers chose which version to use during the observation but were required to transcribe any paper-based data into the online form immediately following the observation and prior to talking with the other observer. During an observation, when an observer experienced activities/behavior that supported UDL alignment (or nonalignment), they underlined the item (or tentatively marked the item in the online version). If they observed what they characterized as more substantiated UDL (e.g., they made a previous judgment that there was preemergent UDL, rating an item as "1") but observed a more active use of UDL within the same context at a subsequent point, they selected the more aligned item on the scale. At the conclusion of the observation, the final, highest rating for each item was circled to represent a final judgment. Fundamentally, the UDL-OMT was designed to be a dynamic observation tool. At any given time within a lesson/activity, it is possible that a UDL guideline/checkpoint may be actively applied (or situationally present within the environment but not applied; see instrument scale; Table 2). The tool was not designed to be an instrument where the observer watched a classroom for an extended period of time and relied on memory to make a large number of judgments only at the conclusion of an observation. Mean scores for each section ([b] through [d])

Table 3. Reliability: Cronbach's Alpha (Internal Stability) and Intraclass Correlation Coefficients for Sections (b) to (d).

UDL area	Cronbach's α^a	Intraclass correlation coefficient (ICC) ^{b,c}
Content representation and delivery (nine items)	.885	.514
Expression of understanding (seven items)	.865	.604
Activity and student engagement (nine items)	.821	.610
Combined (25 items)	.939	.567

Note. UDL = Universal Design for Learning.

^a $\alpha = .8$ to $.899$ = "Good"; $\alpha = .9$ or higher = "Excellent" (Cortina, 1993). ^bICC using average measure ICC; ICC between 0.50 and 0.69 indicates "moderate" agreement (Weir, 2005); ICC between 0.5 and 0.75 indicate "moderate" agreement (Koo & Li, 2016). ^cICC between 0.40 and 0.59 = "fair"; between 0.60 and 0.74 = "good" (Cicchetti, 1994).

were also calculated as well as the overall mean for the completed sections.

Data Analysis

After 22 observations (11 paired classroom observations) occurred, all data were downloaded for analysis. Cronbach's alphas and intraclass correlation coefficients (ICC) were calculated to measure the internal stability/reliability of the instrument with respect to three individual sections of the instrument, as well as the combined items of these sections. ICCs were calculated using a two-way mixed-effect model, designed to identify the degree of absolute agreement and consistency (Koo & Li, 2016). To identify differences in environments, data were also combined and then means scores were calculated to make some general characterizations about the level of UDL implementation within classrooms. These characterizations were then compared with a priori district leadership perceptions by seeking their affirmation with UDL-OMT data.

Results

Reliability of the UDL-OMT

Internal consistency and reliability of the instrument were measured (Table 3). Across all three sections, Cronbach's alphas were above .80, yielding internal consistency considered "Good" (Cortina, 1993). When all UDL items were combined (25 items), the alpha score was above .90, yielding internal consistency considered "Excellent" (Cortina, 1993). All ICC scores were above 0.509, indicating moderate agreement (Koo & Li, 2016; Weir, 2005). According to Cicchetti (1994), ICC scores between 0.40 and 0.59 represent "fair" (Content Representation and Delivery), and ICC scores between 0.60 and 0.74 represent "good" (Expression and Understanding; Activity and Student Engagement).

Characterizing UDL in Classrooms

Three classrooms were identified to have scores between 1.42 and 1.58 (see Table 4). The classrooms that scored in

Table 4. The Degree of UDL Observed Across Classrooms.

Level of UDL use	Average of combined scores (classroom location)	Characterization
Preemergent (1–1.74)	1.42 (elementary school) 1.46 (middle school) 1.58 (elementary school)	UDL was not observed beyond naturally occurring instructional practices.
Emergent (1.75–2.49)	1.98 (elementary school) 2.06 (elementary school) 2.12 (middle school) 2.26 (elementary school) 2.42 (high school)	UDL was observed, but it was not necessarily applied consistently during the observation.
Observed (2.5–3.00)	2.54 (high school) 2.84 (middle school) 2.88 (high school)	UDL was obvious and being consistently applied.

Note. UDL = Universal Design for Learning.

the range of 1 to 1.74 were characterized as having pre-emergent UDL, where UDL was not observed beyond naturally occurring instructional practices (e.g., some students worked in small groups, whereas others worked individually; students appeared to be able to work on a personal computer vs. textbook). Another way of characterizing a preemergent classroom was that, based on the elements of UDL observed, there was not sufficient UDL to support the notion that UDL had been systematically designed for that observation.

Five classrooms had a score between 1.98 and 2.42. Classrooms that scored in the range of 1.75 to 2.49 were characterized as having emergent UDL, where UDL was observed, but not consistently. Another way of characterizing an emergent classroom was that, based on the elements of UDL observed, there were times/events where a UDL checkpoint was noticed, but other times/events where principles/checkpoints of UDL could have been implemented without much difficulty, but were not observed.

Three classrooms had a score between 2.54 and 2.88. Classrooms that scored in the range of 2.5 to 3.24 were characterized as having observed UDL, where the application

of principles of UDL was obvious and consistently applied during observations. All observations were affirmed by district leadership during postobservation debriefing.

Discussion

This initial field test of the UDL-OMT supports the tool's reliability for measuring UDL. Given that the observers purposefully took on different roles in their perspective (i.e., one presuming UDL would be present, with the other more stringent regarding alignment to UDL), Cronbach's alpha scores indicated that there was good to excellent internal consistency. For example, despite one observer being more stringent (and more likely to score items lower), whereas the other observer anticipated UDL would be present (and more likely to score items higher), these observers were internally consistent within their individual observations. Specifically, both observers were consistent in their ratings across the continuum of classrooms, rating the more substantial (i.e., higher scores across UDL-OMT items/sections) alignment to UDL in similar classrooms, while both rating classrooms with less substantial UDL consistently lower.

The primary intention of this field test was to identify the internal consistency of the UDL-OMT. Because of the observers' purposeful difference in their stringency of orientation during the observations, as well as the fact that their attention to specific instructional and learning events was not always in sync and simultaneous (e.g., one observer might have been positioned near and concentrating on a small group of students' working in one part of the room, while the other was in a different location focusing on a conversation between an individual student and teacher), it was unmanageable to accurately measure their interrater reliability, which helps explain why the ICC scores revealed moderate agreement. Observers highly knowledgeable and supportive of UDL may enter into observations being extremely conservative (e.g., "The mere existence of a pen and paper or word processors are not enough to warrant multiple means of expression") or overly liberal (e.g., "If it meets the spirit, it should be called UDL") in their willingness to identify behaviors, strategies, or tools as elements associated with UDL. Future studies focusing on measuring the reliability of the UDL-OMT should be performed on how an observer's background knowledge of UDL, and the degree that expectations regarding the existence of UDL, has an effect on observations. In addition, future studies should investigate interrater reliability to determine whether observers rate the same or different events as UDL and, finally, determine what training may be needed to mitigate any observer bias.

Further Research and Application

Initial field testing of the UDL-OMT demonstrates that the instrument is reliable and enables users to operationalize

their perception of UDL. Observing classroom instruction with the UDL-OMT can help distinguish UDL from other practices. Although the initial need for the tool was identified as a means to support research and evaluation (Basham & Gardner, 2010), the tool is a means to support ongoing observation in environments supporting UDL implementation. With further research, the tool may be able to be used to observe the consistency of UDL implementation (e.g., practitioners and school-based personnel using the UDL-OMT to verify whether a predetermined minimum level of UDL is present), with an eye toward multiple observations to assess continued improvement. UDL is based on proactive and iterative design. The UDL-OMT may likely be more effective when used to measure changes over time rather than as a single-shot summative evaluation.

There are a variety of procedural considerations to be discussed. Like any observation of instruction, conducting multiple observations likely yields more valid information. Considering that the UDL-OMT can be used as a formative tool, practitioners and school-based personnel are free to identify what type of formative analysis works best within the context of desired outcomes. Thus, if the intention is to observe the expansion of UDL in a specific classroom over time, the instrument could be used to make multiple observations. If the intention is to observe current practices for the purpose of providing educator feedback on the nature and degree of UDL implementation, the tool could be used by a group of observers. Each observer makes individual judgments, but then they come together to arrive at consensus across items, thereby providing collaborative feedback based on a uniform instrument.

The context for the environment and timing of the observation is important to consider. For instance, an observer may enter a learning environment on the second or third day of a multiday unit to conduct a single observation (as was the case in some observations conducted as part of this field test). During the time period when the observer is present, students are busy working on gathering and/or manipulating information and beginning to consider various ways they are going to express their understanding. In this case, the observer is less likely to observe and gain a full understanding of the complexities of the environment across the entire spectrum (i.e., multiple means of representation, action and expression, and engagement) of UDL. The application of the UDL-OMT must, therefore, be used with flexibility in mind. If the intention is to assess UDL in a single lesson over the course of 45 min, as long as the measure is only considered valid in the context of the specific lesson (vs. an affirmation of whether UDL is a central framework for a classroom environment), it represents a reliable measure. If the application of the instrument is to determine whether a teacher is implementing UDL, multiple observations should be performed over days, at different times, and during varied activities, thereby providing multiple samples to gauge UDL's implementation.

Fidelity of implementation is a more challenging issue. Unlike the discrete evidence that supports the validity of the specific guidelines or checkpoints within the framework, the UDL-OMT supports a contextualized understanding of how the framework converges into a holistic and dynamic design schema. Such observations could support a further understanding of UDL, including the improvement of the framework itself and better known design archetypes for a variety of learning environments, building upon what Edyburn (2010) calls diversity blueprints.

As this initial field test was focused on basic internal consistency, whether the UDL-OMT can be used as a measure of fidelity requires further research. First, as UDL design is goal driven, any measure of fidelity must associate with the students meeting the goal(s) of designed learning experience. Second, using the UDL-OMT, a teacher could identify target scores across each of the four sections that teachers believe should be present in an observation of their instruction. If a trained observer conducts a series of observations across X instructional sessions and the practitioner maintains or demonstrates increased scores across observations, then it may be concluded that the practitioner has consistency of UDL implementation. Alternatively, if the observations denote a decrease in scores, then it may be a reasonable conclusion that the practitioner has not met fidelity in implementing UDL. Thus, fidelity of implementation for UDL by practitioners may be operationalized differently than fidelity for researchers. Fundamentally, the consideration of how to approach fidelity of implementation as it relates to the UDL-OMT is still in its early phase of discussion as there are currently no established boundaries or parameters that limit the discussion of fidelity as it relates to the UDL-OMT.

Additional Considerations

There remain a number of UDL design, implementation, and observations issues that appear interesting options to pursue. Do observers' viewpoints/movements in classrooms influence their perspective? In this study, the observers were free to move about the room. Might a single static location (as often is the case for classroom observations) yield different perspectives? How silent should an observer be? Might the observance of UDL differ if the observer is free to query a teacher or student about a particular behavior that may or may not be indicative of the inclusion or absence of a UDL guideline? For example, at the conclusion of one observation, the observers were chatting with the teacher, who stated, "I wish you had been here yesterday, we were doing X, Y, and Z." Based on this communication, it was deduced that UDL practices that were not present during the observation had been included in a prior set of student experiences, yet unobserved during our observation. There may be merit to including postobservation procedures, that

enable observers to clarify and/or refine their initial assessment, to reflect their measurement of UDL implementation. Finally, another unresolved issue is whether it is essential that controlled instructional parameters be established, or whether there may be measurement differences based on the length or type of instruction observed. For example, should observations require a minimum time period? Should observations occur (or not occur) during specific activities (e.g., educator-centered instruction, computer-centered instruction, learner-centered instruction) or should an entire lesson be observed from start to finish?

Finally, until additional research is performed, the demarcation points and terms used to characterize levels of UDL use across classrooms (see Table 4) remain subjective. In this initial field test, the difference between the highest classrooms falling into the "Pre-emergent" versus the lowest "Emergent" was relatively clear (a difference of 0.40). The difference between the highest "Emergent" UDL classroom (a score of 2.42) and the lowest "Observed" UDL classroom (a score of 2.54) was 0.12. The three classrooms with scores at or above 2.84 on a scale of 0 to 3 clearly were implementing UDL within the criteria of the UDL-OMT. Whether or not the currently designated demarcation points represent an accurate way to distinguish levels of UDL between classrooms may be discerned through more observations by observers familiar with UDL.

Limitations

As this is an initial field test, there are numerous limitations to consider. A limited number of classrooms were observed. Additional applications, ideally across multiple classrooms and districts, are needed to provide more extensive data that strengthen the statistical analysis to support or refute the instrument's reliability. All observations for this field test occurred in classrooms that were well into their respective instructional activities. Thus, the use of the first section of the instrument—Introducing and Framing New Material—was omitted from the analysis. One of the next research studies using the UDL-OMT should focus on using the UDL-OMT at the onset of a lesson, where the existence or omission of certain instructional strategies and use of guidelines/checkpoints of UDL at the beginning of the lesson may have an effect on student engagement throughout the lesson.

Both observers had extensive knowledge of UDL, including UDL instructional strategies and tools. Although the a priori bias differentiating how the two observers measured UDL did not affect the internal consistency of the UDL-OMT (i.e., meaning the internal consistency of each observer's use of the instrument was not affected), it is unclear whether their knowledge and experiences had an effect of being more internally consistent in measurement.

The level of an observer's understanding and training is a consideration for further research.

Conclusion

As reported in this study, the UDL-OMT has demonstrated reliability when used to measure alignment to UDL in classroom settings. Although initially conceptualized as a research tool, as a result of the experiences in using the tool, one of the more programmatic and pragmatic applications of the UDL-OMT is to use it as an instrument where practitioners and school-based personnel can independently and objectively assess the nature and degree of UDL implementation. The form and function of this process are flexible and can be used to observe commonalities and differences in UDL across activities. The internal consistency of the measure enables observers who may have liberal or more conservative perspectives to remain consistent across different environments or individuals.

In its present form, the UDL-OMT may function well as a formative evaluation tool, where two or more individuals use the instrument as a foundation to discuss and develop a better understanding of UDL implementation in instructional settings. It is also important to note that the circumstances, methods, and procedures used in this study were not designed to evaluate the ability of the tool to measure a school district's fidelity of implementation of UDL, or to make "high stakes" or summative conclusions, such as whether a specific teacher/implementer possesses the appropriate knowledge and skills labeled as a "practitioner" or someone "highly skilled" in implementing UDL. We do not mean to imply that the UDL-OMT cannot be used this way. We simply note that it was never the intention of this initial field test to pursue these types of research questions, and we also acknowledge that testing these aspects will require a different set of a priori conditions (e.g., establishing some degree of standardization regarding observations across implementation environments). In addition, UDL researchers are encouraged to use the UDL Reporting Criteria established to support the reporting of UDL-based interventions (Rao et al., 2018, 2019).

Using UDL-OMT should be through a lens acknowledging the variability of procedural, contextual, and temporal factors. The instructional context (e.g., goals, outcomes, learners, and barriers) associated with learning environments has an impact on the design of the environment and, thus, the implementation of UDL for any single observation. For instance, learning extremely challenging scientific or mathematical content, understanding of key historical events and their societal implications, or writing an acrostic poem relies on different design considerations. As instructional design is focused on learning goals or outcomes (e.g., intellectual skills, cognitive strategies, verbal information, attitudes, or motor skills; Gagne, 1970), they also influence the

observed design of an instructional session. Acknowledgment of these factors helps engage our interest in future research to see how sensitive the UDL-OMT may be when these factors are accounted for. Certainly, these variables are relevant to design factors that should be considered whenever the results of the UDL-OMT are shared with a teacher/instructor.

The UDL-OMT represents a single tool that can be reliably applied to measure UDL implementation in instructional settings. However, measuring UDL continues to be a complex and challenging phenomenon where more research is required. There remain numerous exciting design and procedural UDL measurement-related issues worth pursuing. It is hoped that the UDL-OMT promotes further research and conversation in observing these highly complex learning environments.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

James D. Basham  <https://orcid.org/0000-0003-3667-5168>

References

- Basham, J. D., Blackorby, J., Stahl, S., & Zhang, L. (2018). Universal design for learning: Because students are (the) variable. In K. Kennedy & R. E. Ferdig (Eds.), *Handbook of research on K-12 online and blended learning* (2nd ed., pp. 477–507). ETC Press.
- Basham, J. D., & Gardner, J. E. (2010). Measuring universal design for learning. *Special Education Technology Practice*, 12(2), 15–19.
- Basham, J. D., Koehler, C., & Israel, M. (2011). Creating a "STEM for all" environment. In C. Johnson (Ed.), *Secondary STEM education reform (secondary education in the changing world)* (pp. 1–25). Palgrave Macmillan.
- Basham, J. D., & Marino, M. (2013). Understanding STEM education and supporting students through universal design for learning. *Teaching Exceptional Children*, 45(4), 8–15. <https://doi.org/10.1177/004005991304500401>
- Basham, J. D., Meyer, H., & Perry, E. (2010). The design and application of the digital backpack. *Journal of Research on Technology in Education*, 42, 339–359. <https://doi.org/10.1080/015391523.2010.10782555>
- Basham, J. D., Smith, S. J., & Satter, A. L. (2016). Universal design for learning: Scanning for alignment in K–12 blended and fully online learning materials. *Journal of Special Education Technology*, 31, 147–155. <https://doi.org/10.1177/0162643416660836>

- Bryk, A. S., Gomez, L. M., & Grunow, A. (2011). Getting ideas into action: Building networked improvement communities in education. In M. T. Hallinan (Ed.), *Frontiers in sociology of education* (pp. 127–162). Springer.
- Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2015). *Learning to improve: How America's schools can get better at getting better*. Harvard Education Press.
- CAST. (2011). *Universal design for learning guidelines* (Version 2.0). <http://udlguidelines.cast.org/more/downloads>
- CAST. (2018). *Universal design for learning guidelines* (Version 2.2). <http://udlguidelines.cast.org>
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessments*, 4, 284–290.
- Cobb, P., Confrey, J., DiSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, 32(1), 9–13. <https://doi.org/10.3102/0013189X032001009>
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78(1), 98–104.
- Edyburn, D. L. (2010). Would you recognize universal design for learning if you saw it? Ten propositions for new directions for the second decade of UDL. *Learning Disability Quarterly*, 33(1), 33–41. <https://doi.org/10.1177/073194871003300103>
- Every Student Succeeds Act (ESSA) of 2015, Pub. L. No. 114-95, § 4104 (2015).
- Gagne, R. M. (1970). *The conditions of learning* (2nd ed.). Holt, Rinehart & Winston.
- Hall, T., Strangman, N., & Meyer, A. (2003). *Differentiated instruction and implications for UDL implementation*. National Center on Accessing the General Curriculum.
- Hattie, J. (2011). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Springer.
- Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. Routledge.
- Kennedy, M. J., Thomas, C. N., Meyer, J. P., Alves, K. D., & Lloyd, J. W. (2014). Using evidence-based multimedia to improve vocabulary performance of adolescents with LD: A UDL approach. *Learning Disability Quarterly*, 37(2), 71–86. <https://doi.org/10.1177/0731948713507262>
- King-Sears, M. E. (2009). Universal design for learning: Technology and pedagogy. *Learning Disability Quarterly*, 32, 199–201. <https://doi.org/10.2307/27740372>
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15, 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Lewis, C. (2015). What is improvement science? Do we need it in education? *Educational Researcher*, 44(1), 54–61. <https://doi.org/10.3102/0013189X15570388>
- Marino, M. T. (2009). Understanding how adolescents with reading difficulties utilize technology-based tools. *Exceptionality*, 17, 88–102. <https://doi.org/10.1080/09362830902805848>
- Marino, M. T., Gotch, C. M., Israel, M., Vasquez, E., III, Basham, J. D., & Becht, K. (2014). UDL in the middle school science classroom: Can video games and alternative text heighten engagement and learning for students with learning disabilities? *Learning Disability Quarterly*, 37(2), 87–99. <https://doi.org/10.1177/0731948713503963>
- Meyer, A., Rose, D. H., & Gordon, D. T. (2014). *Universal design for learning: Theory and practice*. CAST Professional.
- National Center on Universal Design for Learning. (2012). *UDL implementation: A process of change* [Online seminar presentation] (UDL Series, No. 3). <http://udlseries.udlcenter.org/categories/implement.html>
- National Educational Technology Plan. (2017). *Reimagining the role of technology in education*. <https://tech.ed.gov/netp/>
- Ok, M. W., Rao, K., Bryant, B. R., & McDougall, D. (2017). Universal design for learning in Pre-K to Grade 12 classrooms: A systematic review of research. *Exceptionality*, 25, 116–138. <https://doi.org/10.1080/09362835.2016.1196450>
- Piburn, M., & Sawada, D. (2000). *Reformed Teaching Observation Protocol (RTOP) Reference manual* [Technical report]. Arizona Collaborative for Excellence in the Preparation of Teachers.
- Proctor, C. P., Dalton, B., Uccelli, P., Biancarosa, G., Mo, E., Snow, C., & Neugebauer, S. (2011). Improving comprehension online: Effects of deep vocabulary instruction with bilingual and monolingual fifth graders. *Reading and Writing*, 24, 517–544. <https://doi.org/10.1007/s11145-009-9218-2>
- Rao, K., Ok, M. W., & Bryant, B. R. (2014). A review of research on Universal Design Educational models. *Remedial and Special Education*, 35, 153–166. <https://doi.org/10.1177/0741932513518980>
- Rao, K., Ok, M. W., Smith, S. J., Evmenova, A. S., & Edyburn, D. (2019). Validation of the UDL reporting criteria with extant UDL research. *Remedial and Special Education*. <https://doi.org/10.1177/0741932519847755>
- Rao, K., Smith, S. J., Edyburn, D., Grima-Farrell, C., Van Horn, G., & Yalom-Chamowitz, S. (2018). *UDL reporting criteria* [Developed by a working group of the Universal Design for Learning Implementation and Research (UDL-IRN) Research Committee]. <https://udl-irn.org/udl-reporting-criteria/>
- Rappolt-Schlichtmann, G., Daley, S. G., Lim, S., Lapinski, S., Robinson, K. H., & Johnson, M. (2013). Universal Design for Learning and elementary school science: Exploring the efficacy, use, and perceptions of a web-based science notebook. *Journal of Educational Psychology*, 105, 1210–1225. <https://doi.org/10.1007/s11145-009-9218-2>
- Rose, D., & Meyer, A. (2000). *The future is in the margins: The role of technology and disability in educational reform* (ERIC Document ED451624). <https://files.eric.ed.gov/full-text/ED451624.pdf>
- Rose, D., & Meyer, A. (2002). *Teaching every student in the digital age: Universal design for learning*. Association for Supervision and Curriculum Development.
- Schutt, R. K. (2018). *Investigating the social world: The process and practice of research*. SAGE.
- Shah, N. (2012, May 23). Study finds districts buying into universal design: But many education leaders still aren't sure what UDL is. *Education Week*, 31(32), 7. <https://www.edweek.org/ew/articles/2012/05/23/32udl-2.h31.html>
- Smith, S. J., Rao, K., Lowrey, K. A., Gardner, J. E., Moore, E., Coy, K., . . . Wojcik, B. (2019). Recommendations for a national research agenda in UDL: Outcomes from the UDL-IRN preconference on research. *Journal of Disability Policy Studies*, 30, 174–185. <https://doi.org/1044207319826219>

- Tokuhamma-Espinosa, T. (2010). *Mind, brain, and education science: A comprehensive guide to the new brain-based teaching*. W.W. Norton.
- Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of Strength and Conditioning Research*, 19, 231–240.
- Wolfram Alpha. (2018). <https://www.wolframalpha.com/>
- Zhang, L., Basham, J. D., & Lowrey, K. A. (2020). Foundations for reinventing the global education system: Personalized learning supported through universal design for learning. In T. D. Neimann & U. M. Stelson (Eds.), *Challenges and opportunities in global approaches to education* (pp. 146–163). IGI Global.