LLT Journal, e-ISSN 2579-9533, p-ISSN 1410-7201, Vol. 24, No. 1, October 2021

LLT Journal: A Journal on Language and Language Teaching
http://e-journal.usd.ac.id/index.php/LLT
Sanata Dharma University, Yogyakarta, Indonesia

INCORPORATING INCLUSIVE ASSESSMENT PRINCIPLES AND UNIVERSAL DESIGN FOR LEARNING IN ASSESSING MULTICULTURAL CLASSROOM: AN AUTOETHNOGRAPHY

Sisilia Kusumaningsih
The University of Montana, The United States
correspondence: sisilia.kusumaningsih@umontana.edu
DOI: 10.24071/llt.v24i1.3656
Received 3 September 2021; accepted 20 October 2021

Abstract
Teaching a multicultural classroom has always been challenging for educators. With the rush toward global standardization, there is a concern that uniformity in education will sacrifice students' freedom in expressing their skills and knowledge. This problem can be seen in the way students are being assessed. The rapid growth of standardized tests and the demand to meet yearly progress thresholds are tangible examples of how assessments may limit the way students demonstrate their learning attainment. This article aims at presenting several alternatives teachers could take to implement Universal Design for Learning (UDL) and inclusive assessment principles in assessing multicultural classrooms. This analytic autoethnography study yielded three themes: providing options of assessment forms, providing accommodation in conducting assessments, and using the strength-based language in assessment feedback.

Keywords: universal design for learning, inclusive assessment, multicultural classrooms

Introduction
The emergence of globalization in the last two decades has shifted the way societies perceive the dissemination of knowledge. Instead of celebrating the uniqueness of each student's culture and integrating it into education, the current education system attempts to create uniformity or homogeneity. Educators strive to meet the global standardization that serves as a benchmark for quality education throughout the world. There has been a rampant spread of standardized tests to assess students' competencies in a particular subject. Consequently, "schools serving multicultural students are often compelled to narrow the curriculum even further to boost test scores" (Volante, 2008).

The widespread use of such tests provokes a continuing debate among educators and raises the ultimate question of how to boost students' scores without sacrificing multiculturalism in classrooms. It is fundamental for teachers to revisit the ultimate goal of multicultural education when determining how to assess learning. Multicultural education should encompass creating equity for students from diverse cultures (Nieto, 2012). Teachers may set one common goal

in each class instruction but may not limit the ways students achieve the goal. Multicultural education opposes the idea of a 'one-size-fits-all' system.

Multicultural education needs inclusive assessment practices, which rely on both quantitative (e.g., test scores) and qualitative data (teacher's notes on student progress, feedback from teammates when students do group work) (Cuseo, 2015). "Student achievement, whether defined as fixed scores or progress over time, is a way to estimate educational effectiveness" (Baker, 2013, p. 88). Student assessment gives a wide perspective of how effective the learning process is for both the students and the instructors. Through assessments, instructors are able to monitor students' growth and evaluate their teaching approaches. "Inclusive assessment and feedback processes are essential if the diversity of our students is to be recognized, value, and supported" (Morris et al., 2019, p.445). However, providing assessments that yield qualitative data in multicultural classrooms has become more challenging when the yearly progress thresholds are indicated by students' standardized test scores. It is undeniable that learning success frequently is measured by quantitative data only, assuming that students with higher scores have a better mastery of learning materials. This paper discusses several alternatives of assessments and feedback processes for diverse classrooms, especially in language teaching contexts. Tangible practices in providing assessment based on Universal Design for Learning (UDL) and inclusive assessment principles are discussed.

What is Universal Design for Learning?

Universal Design for Learning (UDL) emerged from the architectural concept of universal design initiated by Ronald Mace in 1998. Universal design was initially intended to address individuals with disabilities and focused on designing accessible spaces to meet the Americans with Disabilities Act (ADA) requirements. In the current learning context, UDL advocates inclusive learning and encourages teachers to provide different learning modalities for all students. One illustration to help us understand how UDL works is probably by imagining the function of a ramp. A ramp is designed to ease all people with or without disabilities to access or navigate through areas of different heights. A man in a wheelchair, a pregnant woman, a muscular guy carrying heavy stuff, and a child who loves running can all enjoy the benefit of a ramp. In other words, just like a ramp, UDL is intended to ease all individuals in achieving their learning goals. It is designed to facilitate all learners, despite their ability, disability, age, gender, cultural, and linguistic background (CAST, 2018).

In addition, CAST (2018) details the fundamental principles of UDL, which include multiple means of engagement, multiple means of representation, and multiple means of action and expression. In line with the characteristics of performance-based assessments, UDL intends to provide students with numerous options in gaining knowledge and demonstrating their knowledge attainment. One example of UDL implementation in assessment is providing students with read-aloud tools when engaging students in Computer-Assisted Testing (CAT) to minimize the barriers for students with vision impairment and reading disabilities. Another example of the implementation of universal design for learning is providing students with two or more options of assessment formats. In assessing English Language Learners' (ELLs) ability to understand the plot of a literary

work, for example, a teacher might assign students to submit a written book report for those who love writing or a podcast for those who prefer to speak. Other students might choose to draw and record their voices when they explain their drawings. "This flexibility provides relevance and contextualizes the information, which are both hallmarks of supporting culturally diverse students" (Orosco & Klingner, in Rao &Torres, 2017).

What Are the Principles of Inclusive Assessment?

Welch (2000) interestingly explains the concept of inclusive assessment by stating: "it would be unfair to keep glasses from those who need them and equally unfair to make everyone wear glasses; it is my job, as a teacher, to make sure everyone gets the help they need and that help will be different for each student" (p. 38). This notion is similar to the key principle of UDL. Both UDL and inclusive education respect the students' strengths and different needs. Inclusive assessment encourages the use of "different evaluation methods to assess student achievement." (Cuseo, 2015, p. 2). Implementing inclusive assessment principles is deemed practical to improve both validity and equity of evaluation because it provides balanced numerical and non-numerical data.

The use of numerous evaluation methods in inclusive assessment enhances the validity of assessments in a way that the limitation of one test type is more likely to be compensated by another evaluation system (Cuseo, 2015). Similarly, inclusive assessment improves equity because it allows students to demonstrate their ability through different assessment formats to accommodate their various skills and learning styles. Students who have an anxiety disorder, for example, might feel higher pressure when they have to work on a set of norm-referenced tests within a particular time limit. Consequently, they might not perform well. Kaur (2006) mentioned that one of her respondents stated: "I don't like the discomfort and anxiety I have to face in the class during exam and evaluation time" when responding to the interview about norm-referenced test implementation. Those students might enjoy the benefit of having a take-home assignment and thus, perform much better.

Educational Policies Committee and the Affirmative Action/Cultural Diversity Committee (1995) propose that a multicultural education considers three factors that affect students' learning style: socialization process, sociocultural tightness, and ecological adaptation. An culturally inclusive assessment thus, should consider how characteristics of the society, the condition of the environment where students live, and the learning styles are bonded. For example, students who live within a community with a strong social bond and highly valued collectivism will enjoy the activities and be assessed in groups. Conversely, students from the community with less tight social bonds prefer to have an individual assessment. Hence, providing numerous forms of individual and group assessment in a multicultural classroom is fundamental to accommodate the students' characteristics.

Why Do We Need More Than Just Norm-Referenced Tests in Multicultural Classrooms?

Banks (2013) believes that multicultural education is "education for freedom" (p. 9). Education must be perceived as an effort to enrich every individual.

However, studies showed that the widespread use of standardized tests at schools pushes teachers to implement test-based curriculum enactment that limits the flexibility of the instruction and assessment (Murral, 2002; Darling-Hammond & Wise, 1989). "Many teachers feel that they have to sacrifice student creativity and interests by teaching test-taking strategies within a test-oriented curriculum" (Bush, 2005, p. 334). In brief, taking a shortcut by teaching and learning to prepare for the tests means taking the freedom out of learning.

Allman et al. (2004) argued that the purpose of standardized tests, which are norm-referenced, is to compare "the students' performance on the test to the performance of other similar students who have taken the same test (the norm group)" (p. 3). The data obtained from a norm-referenced test is intended to inform how a particular student performs in comparison to the norm group or other similar test-takers rather than to inform a student's learning. While norm-referenced tests could be a valuable tool to "diagnose students' weaknesses in academic performance, they are not useful in determining a student's overall academic capacity in student learning, critical thinking, and higher-order reasoning skills." (Bush, 2005, p.332).

More specifically, norm-referenced tests are designed to satisfy the discrimination index, meaning half of the total test takers should answer the questions correctly, while the other half answer the questions incorrectly. As stated by Salkind (2017), the group that responded to the question correctly is in the upper half that took the test, which eventually results in a discrimination index of 1.00, and the norm group scores "will fall perfectly into a bell curve" (p.105). Therefore, it is worthy to note that norm-referenced assessment might include questions that are not part of the materials. The congruence of learning materials and questions is not the primary consideration in this type of test. It creates more barriers for students because when there is a discrepancy between assessment and the materials delivered in the instruction, students' success in learning is unlikely (Segers et al., 2001)

On the other hand, it is undeniable that teachers "may put their professional reputation and status at risk because of punitive sanctions they can experience if the test scores of their students do not increase between testing cycles" (Banks, 2013, p.9). For example, in the United States, schools are required to meet the Adequate Yearly Progress (AYP) criteria. In the United States, public schools that do not meet the AYP benchmark, "as reflected in mandated improvements in test scores, are labeled as failing and are eventually taken over by the state" (Volante, 2008, p.12.). This issue created a dilemma between meeting the benchmark criteria and focusing on different ways students learn. Despite the dilemma, Allman et al. (2006) remind teachers to distinguish between coaching and cheating in preparing students to succeed in assessments. A teacher who attempts to coach the students will focus on teaching the content and skills covered in the tests. They will provide different types of assessments to prepare students for success and ensure that they have qualitative and quantitative data to determine students' progress. On the other hand, cheating focuses on the test-based curriculum enactment by implementing a drilling method and asking students to do the tests simulation repeatedly. While it is true that the success in standardized tests is arguably because of the familiarity with the test format over time (Linn, 2000; Popham, 2001), test-based teaching sacrifices the culturally responsive

practice. Despite their different cultural/linguistic backgrounds and learning strengths and weaknesses, students must do the same assessment format.

Why Do We Need Performance-Based Assessments?

Coelho (2012), in his study about language and learning in multicultural classrooms, found that performance-based assessments create more inclusive education. This assessment is open-ended, and thereby, provides more opportunities for students to incorporate their prior knowledge and skills in various ways. More specifically, performance-based assessment demands students to be able to justify their answers. Therefore, it requires a high-quality product or performance where correctness is not the only criterion measured (Center for Innovative Teaching and Learning, 2019). As stated by Ambrosetti and Cho (2005), performance-based assessments "place the responsibility of learning on the students and provide the opportunity for them to engage in activities that embrace their various learning styles and multiple intelligences" (p. 59). Tanner (2001) believed that performance-based assessment helps teachers to be more aware of the interconnectivity between students backgrounds and what happen in the classroom, so that "some of the disadvantages that culture and language minority students must shoulder" can be neutralized (p. 28). In providing performance-based tasks, teachers need to be culturally aware and guide the students to "compare, relate, or apply what they know to new information and experiences" (Goodwin, 2000, p.7). Hence, the performance-based assessments will be more accommodative than traditional assessments that require students from diverse backgrounds to work on the same tests.

Gielen et al. (2003) proposed two important reasons for implementing performance-based assessment: construct validity and consequential validity. Construct validity refers to the degree an assessment measures what it is supposed to measure, while consequential validity describes the effects of assessment on instruction and student learning. Both construct validity and consequential validity are measured under specific criteria. Construct validity is demonstrated in the appropriateness of a task to reflect the competencies that need to be assessed, the content of the task that represents a real-life problem of the knowledge domain assessed, and the resemblance of thinking process in real-life problem solving (Gulikers et al., 2004). Based on those principles, it is believed that performance-based assessments have higher construct validity for measuring competencies than norm-referenced assessments.

The second reason for implementing performance-based assessment is its consequential validity, which can be seen through the alignment of instruction, learning, and assessment (Gulikers et al., 2004). If students are to evaluate a descriptive paragraph, for example, a performance-based assessment will most likely ask students to read an authentic brochure from a zoo that describes the animals living there. They will then be asked to demonstrate their understanding by creating a mind map that summarizes the information in the paragraphs before writing their own descriptive paragraph. This example shows how performance-based assessment is more open-ended and more aligned with the input. Rather than asking students to answer multiple-choice questions about a descriptive paragraph and what features need to be in the paragraph, performance-based assessment requires students to engage in paragraph writing.

Allman et al. (2004) assert that "performance-based assessment requires students to produce a product or demonstrate a process, solve a problem involving several steps, or carry out an activity that demonstrates proficiency with a complex skill" (p. 38). This type of test offers an open-ended model and forms that allow students to work both individually and in groups. The assessment may vary, from a personal essay, an in-pair presentation, a group discussion to a community project. Furthermore, Allman et al. (2004) recommend selecting performance-based tasks that correspond to the real environment outside the classrooms. In brief, there are two main principles in creating and conducting a performance-based assessment that must be considered, which include the authenticity of the task and the variety of the assessment forms.

Method

This inquiry is in the form of an analytic autoethnography study. Autoethnography can be defined as an approach that "describes and systematically analyzes (graphy)" the researcher's "personal experience (auto) to understand cultural experience (ethno)" (Ellis et al., 2011, p. 273). In other words, this approach puts a researcher at the center of an inquiry. As a research method, autoethnography combines the tenets of autobiography and ethnography (Ellis et al., 2011)

Due to its nature that is centered on the self, autoethnography oftentimes is considered as "uncontrolled, subjective and anecdotal" (Clark & Gruba, 2010, p. 166). Despite the criticism, Etherington (as cited in Clark & Gruba, 2010) believes that this approach value and "legitimate the inclusion of the researchers' self and culture as an ethical and politically sound approach" (p. 166). To make the analysis more objective and specific, other scholars introduced analytic autoethnography, wherein researchers present theories and former research findings to support the reflection (Anderson, in Canagarajah, 2012). Based on that notion, this study does not only present the practices in implementing UDL and inclusive assessment principles but also links those tangible practices to the former studies.

Findings and Discussion

Reflecting upon my experiences as both a student and a teaching assistant at one of the universities in the Rocky Mountain area, three themes emerged in the analysis: providing options of assessment forms, providing accommodation in conducting assessments, and using strength-based language in assessment feedback.

Providing Options of Assessment Forms

The first theme to emerge in the study was providing options, which is mainly based on the UDL key learning principle. By giving options, teachers gave students opportunities to submit their assignments/ tasks for feedback in multiple representations. Echoing the third principle of UDL, the practice of providing choices in assessment requires multiple media of communication and different types of technology. One example that stood out the most in my reflection is the option to turn in the assignment in numerous forms (e.g., storybook designed through story jumper, podcast/ted talk, poster, etc.). Reflecting on my experiences

of using story jumper, I believe this app is an excellent tool to teach literary works in the target language. Students will have an opportunity to integrate images, texts, and sound in one storybook. Allowing students to work with such kind of application is beneficial because it provides both multi modalities and scaffolding for students. Students with language impairment, for example, might enjoy the benefit of the different options of illustration to help them express the ideas in written forms. More specifically, Cowan and Cress (in Ezeh, 2021) posit that "considering the importance of giving the second language (L2) learners diverse opportunities to create and represent meaning that support multimodality are critical for 21st century linguistically and culturally diverse classrooms" (p. 1). Additionally, the bookmaking application, such as story jumper, provides students different modes to present their works and bridges language skills and digital literacy (Ezeh, 2021).

In addition to incorporating technologies, it is suggested that teachers include cultural content into the assessment criteria. For example, the story that is presented must demonstrate social justice or inclusivism. Teachers might also require that the literary works written must cover some arguments or thoughts from minority groups. This type of assessment allows students to explore multicultural issues. As explained by Nieto (2012), one of the levels of multicultural education is promoting respect by exposing students to "different ways of approaching the same reality" so that "they would expand their way of looking at the world." (p. 394). By providing options for students to present their works, a teacher has demonstrated how a goal can be approached in numerous ways and how an issue can be addressed from different point of views.

Another example of performance-based assessment at the [name of the university] that incorporates the UDL principles is the use of comprehension constructors and double-entry journals in reading assessment. Instead of asking students to work on multiple-choice questions, this reading assessment allows students to express their understanding of the texts in charts. In their summary, students write 'so what?', which will then guide them to conclude in what way the content of the reading is beneficial for them. As mentioned by Tovani (2004), "designing comprehension constructors is a constant process of matching content, readers, and goals" (p.78). This constructor serves as a practical formative assessment that facilitates readers in connecting the different ideas in the text (e.g., how one idea in a text will support other main ideas). Unlike norm-referenced test that is more rigid and requires one correct answer, this type of assessment allows teachers to see students' various perspectives in interpreting texts. Oftentimes, different cultural and linguistic backgrounds affect the readers' interpretation of texts. The comprehension of a text will depend on readers' "individual perspectives and background -because meaning and structure of a text are not inherent in the print but are invited by the author and imputed to the text by the reader" (Liu, 2014, p.1089).

Similar to a comprehension constructor, a double-entry journal requires students to compare and contrast the ideas discussed in texts in two different columns. This assessment will be suitable for assessing students' ability to understand argumentative essays or to compare and contrast two different types of texts. Double entry journal emphasizes the importance of creating connections between text-to-text, text-to-self, and text-to-the-world (Tovani, 2004). In line

with the principle of performance-based assessment, particularly the authenticity principle, double entry journal demands students to link the text to the issues in a real-life context. When analyzing a text about Dakota Pipeline issue, for example, students were asked to outline the points that justify both Native Americans' perspective and the government stands. Students should collect additional information from other articles and list the reasons to justify the people's decision to refuse the pipeline construction. They were then required to list the reasons to justify government's decision in building the pipeline access. At the end of the evaluation, students could think-pair-and share their ideas and brainstorm the possible solution to mediate the issue. Both comprehension constructor and double-entry journal manifest the UDL principle to support executive function (e.g., helping students set goals, self-regulate, and develop strategies) (Rao & Torres, 2017).

The strength of providing more options and more performance-based assessment is that it will give students more autonomy in learning. Students enjoy the benefits of having the freedom to express their understanding. However, this type of assessment is more prone to teachers' subjectivity. Unlike multiple-choice questions, which ease teachers in determining the correct or incorrect answer, performance-based assessments do not explicitly show what is right or wrong. It reflects the depth of students' understanding of a particular lesson unit and shows how each student interprets the knowledge they attain. Therefore, it is imperative to prepare rubrics that inform teachers about the detailed criteria and aspects being assessed. Being clear on what is expected from each assessment will be the key to the success of the performance-based assessment.

Provide Accommodation in Conducting Assessment

The second theme that emerged in the study was to provide accommodation in conducting assessments. While performance-based assessments provide more information on students' attainment than norm-referenced tests, we understand that the implementation of b oth tests is not an either-or situation. Both tests serve different purposes, and teachers need to assess students with different types of tests. When implementing both types of assessments, especially the norm-referenced one, teachers could provide accommodation to support students from diverse cultures and linguistic backgrounds. Providing accommodation means making adjustments to how students are taught and how they will be assessed. Allman et al. (2004) define accommodations as the adaptation made "to the way students are instructed and how they are tested." Accommodation and modification are frequently used interchangeably. However, these two terms serve different meanings. Unlike modification that changes what students learn and what is being assessed, accommodation expects students that need support (e.g., students with disabilities, ELL learners in English speaking countries) to master the same learning content and achieve the same objectives as other students. Sometimes teachers hesitate to provide accommodation because of the assumption that providing accommodation means underestimating students' competence. It is important to note that by providing accommodation, we still set the bar high for all students, but we need to adjust the way we help them achieve the goal.

Some forms of accommodations implemented at [name of the university] and at my workplace are giving additional time for students to accomplish the task,

including paraprofessional/peer partner/tutor to assist them in finishing the assignments, and simplifying test instruction. For example, students who have dyslexia in my classroom would be given extra time to work on their reading tests. They also had an opportunity to work one-on-one with a tutor when they work on their assignment.

The accommodations provided in my classroom were in the form of direct and indirect linguistics accommodation. Direct linguistic support refers to adjusting the language complexity of the assessment, which can be rarely implemented when the construct being tested is the language proficiency itself Rivera and Collum (2004). A tangible practice of direct linguistic accommodation for ELLs is a linguistic simplification, which encourages concise and straightforward language/word choices in test items/test directions without altering the construct being assessed (Rivera & Collum, 2004). Providing the help from tutor/paraprofessional is another example of direct linguistic supports. Meanwhile, indirect linguistic support refers to the adjustment of the environment wherein the assessment is conducted. Giving additional time belong to the indirect linguistic support.

The strength of accommodation is that the assessment will be designed and adjusted to the students' needs and conditions. Parents, students, and paraprofessionals must be involved when teachers are to design the appropriate accommodation. One of the challenges in providing accommodation is the time constrain. Arranging a plan and determining the time that will work for all parties are not easy tasks. Another challenge might come from the students' families. Jenkins (1969) argues that parents have different past school experiences, relationships with the children, and values and goals. Those aspects may affect their attitude towards schools and their perspectives on the importance of school-family relationships. Some families might prefer not to be involved in the meetings to design the accommodation for the students

The Use of Strength-Based Language in Assessment Feedback

Another theme yielded from the reflection is the implementation of strength-based language in self-assessment and peer feedback. UDL encourages teachers to provide options for feedback. Alternating between self-assessment, peer feedback and teacher feedback will be one of the manifestations of this principle. To make sure that the feedbacks given are respectful and inclusive, it is necessary to use the strength-based language. One example of using strength-based language in providing feedback is by implementing tootling in my classes. The basic principle of tootling is focusing on the positive aspects of every action or behavior and expressing the concerns towards the negative ones in a sympathetic manner. Many scholars conducted studies on the implementation of tootling in inclusive classrooms and proved that tootling impacted students' disruptive behavior decline.

Cihak et al. (2009) describe tootling as "a term that was constructed from the word 'tattling' and the expression 'tooting your own horn." As opposed to tattling, tootling encourages students to observe their peers' actions and prosocial behaviors and express concerns in a sympathetic way. Adapting tootling to assess students' knowledge attainment might be one alternative to engage students in giving constructive feedback to their peers. To implement tootling, I provided one

box for each student to place all tootles received from their classmates and teachers. Assessment and tootles might be given at the end of every unit of the lesson. I asked students to either summarize the main point of the unit's content or write an essay to demonstrate their understanding of the topic discussed in the unit. Students then have to express their opinion on the materials' content and connect the important points from the lesson with their daily life contexts. For example, students were asked to describe a public figure that inspired their lives upon learning a lesson about describing people.' In addition to write a description, they have to provide criticism on the materials. For example, a student may write: "while mentioning someone's religion in a descriptive text is a common thing in Indonesia, it is considered as unnecessary or probably impolite in another culture."

Upon writing a description and criticism, every student must read one or two works of their peers and write tootles in given cards. Each tootle must elaborate the strength of each writing and some constructive comments on the aspects that need improvements. We need to help our students create a counter-question whenever they make a judgmental comment without considering the culture and perspective of others. Instead of writing, "It is unnecessary to mention the character's religion in the descriptive text," students are expected to write, "I noticed that you discuss the character's religion in your description in an interesting way. In my culture, religion is considered as something personal and we rarely discussed it. Thus, I have some questions regarding the……." By doing so, students learn not only from textbooks or the teacher but also from their peers. Students will also have an opportunity to share the knowledge of the language aspects and deepen their understandings of their own culture by seeing it from the perspective of others.

One challenge that teachers might face is to ensure that all students provide appropriate feedback and criticism towards their peers' works. Since one student might receive two or more tootles from their classmates, there is a probability that one classmates' suggestions or feedbacks contradict the other classmates' comments. To overcome this problem, teachers could provide feedback that summarizes all tootles given to every student to help students conclude the information they received from their peers. Furthermore, it is essential to discuss the points that need some improvements in tootling. Even though providing feedback and summarizing tootles might be time-consuming, it will help teachers learn from students and understand them better.

Conclusion

Assessing a multicultural classroom is a complex process because teachers must ensure that the tasks mirror students' uniqueness. Implementing UDL principles and using an inclusive assessment lens could be helpful for teachers when assessing multicultural classrooms. Overall, this autoethnography yielded three key themes that teachers should consider in providing UDL-based and inclusive assessments. Those three themes include providing options of assessment forms, providing accommodations in conducting assessments, and using strength-based language in assessment feedback. These three themes lead to a conclusion that teachers could advocate the inclusive assessment practice by

reducing the barriers for culturally and linguistically diverse students and adjusting the environment wherein students undergo the assessment.

References

Allman, C., Beech, M. & Thomas, J. (2004). *Assessment for the diverse classroom*. Florida: Bureau of Exceptional Education and Student Services.

Ambrosseti, D. & Cho, G. (2005). Synergism in learning: A critical reflection of authentic assessment. *The High School Journal*, *89*(1), 57-62. doi: 10.1353/hsj.2005.0015

Baker, E. (2013). Critical moments in research and use of assessment. *Theory Into Practice, 52*(Sup1), 83-92. doi: 10.1080/00405841.2013.795445

Banks, J. A. (2013). For freedom's sake. *Transforming curriculum for a culturally diverse society*, *75*.

Canagarajah, A. S. (2012). Teacher development in a global profession: An autoethnography. *Tesol Quarterly*, *46*(2), 258-279. doi: 10.1002/tesq.18

Center for Applied Special Technology. (2018). *Universal design for learning guidelines version 2.2*. Retrieved on August 10, 2021, from http://udlguidelines.cast.org

Center for Innovative Teaching and Learning. (n.d). Retrieved on October 28, 2019, from https://citl.indiana.edu/teaching-resources/assessing-student-learning/authentic-assessment/

Cihak, D. F., Kirk, E. R., & Boon, R. T. (2009). Effects of classwide positive peer "tootling" to reduce the disruptive classroom behaviors of elementary students with and without disabilities. *Journal of Behavioral Education*, *18*, 267. doi: 10.1007/s10864-009-9091-8

Clark, C. & Gruba, P. (2010). The use of social networking sites for foreign language learning: An autoethnographic study of Livemocha. In C.H. Steel, M.J. Keppell, P. Gerbic & S. Housego (Eds.), Curriculum, technology & transformation for an unknown future. *Proceedings Ascilite Sydney 2010* (pp.164-173). Retrieved on July 29, 2021, from http://ascilite.org.au/conferences/sydney10/procs/Cclark-full.pdf

Coelho, E. (2012). *Language and learning in multilingual classrooms*. New York: Multilingual Matters.

Cuseo, J. (October, 2015). *Effective culturally-inclusive assessment of student learning and academic performance*. Retrieved on August 28, 2021 from https://www.researchgate.net/publication/282709203_Effective_Culturally-Inclusive_Assessment_of_Student_Learning_Academic_Performance

Ellis, C., Adams, T. E., & Bochner, A. P. (2011). Autoethnography: An overview. *Historical social research/Historische sozialforschung*, 273-290. Retrieved on August 24, 2021, from https://www.jstor.org/stable/23032294?seq=1&cid=pdf-reference#references_tab_contents

Ezeh, C. (2021). A comparison of storyjumper with book creator, and storybird for multimodal storytelling. *TESL-EJ, 24*(1), 1-9. Retrieved on August 2, 2021, from https://ssrn.com/abstract=3802725

Gielen, S., Dochy, F., & Dierick, S. (2003). The influence of assessment on learning. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimising new*

modes of assessment: In search of quality and standards (pp. 37–54). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Goodwin, L.A. (2000). *Honoring of knowing*. Columbia University.

Gulikers, J. T., Bastiaens, T. J., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessments. *Educational Technology Research and Development*, *52*(3), 67-86. Retrieved on August 24, 2021, from https://link.springer.com/article/10.1007/BF02504676

Kaur, A., Noman, M., & Nordin, H. (2017). Inclusive assessment for linguistically diverse learners in higher education. *Assessment & Evaluation in Higher Education, 42*(5), 756-771. doi: 10.1080/02602938.2016.1187250

Linn, R. L. (2000). Assessment and accountability. *Educational Researcher, 29*, 4-16. doi: https://doi.org/10.3102%2F0013189X029002004

Liu, S. (2014). L2 reading comprehension: Exclusively L2 competence or different competences? *Journal of Language Teaching & Research, 5*(5). doi:10.4304/jltr.5.5.1085-1091

Morris, C., Milton, E., & Goldstone, R. (2019). Case study: Suggesting choice: inclusive assessment processes. *Higher Education Pedagogies, 4*(1), 435-447. doi: https://doi.org/10.1080/23752696.2019.1669479

Nieto, S. (2012). *Affirming diversity: The sociopolitical context of multicultural education*. Boston, MA: Pearson Education

Popham, W. J. (2001). *The truth about testin*g. Alexandria, VA: Association for Supervision & Curriculum Development.

Rao, K., & Torres, C. (2017). Supporting academic and affective learning processes for English language learners with universal design for learning. Tesol QUARTERLY, 51(2), 460-472. Retrieved on August 20, 2021, from https://www.jstor.org/stable/44984766

Rivera, C., & Collum, E. (2004). *An analysis of state assessment policies addressing the accommodation of English language learners*. National Assessment Governing Board.

Salkind, N. J. (2017). *Exploring research*. Boston, MA: Pearson

Segers, M., Dierick, S., & Dochy, F. (2001). Quality standards for new modes of assessment. An exploratory study of the consequential validity of the overall test. *European Journal of Psychology of Education*, *16*(4), 569–586. Retrieved on July 30, 2021, from https://link.springer.com/article/10.1007/BF03173198

Tanner, D. E. (2001). Authentic assessment: A solution, or part of the problem? *The High School Journal, 85*(1), 24-29. Retrieved on January 9, 2021, from https://www.jstor.org/stable/40364377

Tovani, C. (2004). *Do I really have to teach reading?: Content comprehension, grades 6-12*. Portland, Me: Stenhouse Publishers.

Volante, L. (2008). Equity in multicultural student assessment. *The Journal of Educational Thoughts, 42*(1), 11-26. Retrieved on on January 9, 2021, from http://www.jstor.com/stable/23765469