

# Development and Validation of a Brief Digital Pedagogy Competency Scale (DiPeCoS)

**Shraddha RAWAT, Shreya TIWARI, Mayank SHARMA & Nandini CHATTERJEE\***

*UNESCO Mahatma Gandhi Institute of Education for Peace and Sustainable Development,  
35, Ferozshah Road, New Delhi – 110 001, India*

*\*[n.chatterjee@unesco.org](mailto:n.chatterjee@unesco.org)*

**Abstract:** Digital pedagogy or the purposeful application of digital technology in teaching and learning, has the potential to significantly enhance learning experiences. Considering this potential and the rapid digitization in education, it has become critical to define, measure and build teacher's digital pedagogy competence. Although there are a number of self-report tools to evaluate digital pedagogy competence, there is a paucity of scenario-based tools for the same. Scenario-based assessments allow demonstration of knowledge and skills as well as its real-world application. We present here, DiPeCoS, a short tool that assesses a teacher's digital pedagogy competence through their decision making under teaching learning scenarios. Using an item response analysis, a pool of 10 items was reduced to an 8-item scale and validated on a sample of 1315 teachers in India. The DiPeCoS demonstrated unidimensionality, and its constituent items showed acceptable levels of discrimination, difficulty and guessing parameters. Additionally, it also demonstrated acceptable values of reliability. We hope that this scale can be used to conduct training needs assessment amongst teachers and to assess the effectiveness of digital pedagogy training in future research.

**Keywords:** Digital Pedagogy, India, Teacher Competence Scale, Universal Design for Learning, Teacher Digital Competence

## 1. Introduction

The rapid advancement in technology in the 21st century presents a landscape shaped by technology that places new demands on the teaching learning experience and has broad implications both in terms of what it means to teach, and how one teaches and creates a learning experience. Present day teachers do not just need technological knowledge and ICT skills but also the ability to leverage technology affordances for modernizing teaching practice. Numerous efforts have been made to promote technology-enabled learning through ICT training for teachers. This approach has not considered the fact that the intentional use of technology affordances by teachers to achieve learning goals requires a refreshed set of digital pedagogy competence (Fernández-Batanero et al., 2021, Sailer et al., 2021).

Merely integrating technology in educational programs with the aim of 'building digital skills' does not fully utilize the power of meaningful application of technology to transform the learning experience. Research suggests that 'learning occurs when access to technology is combined with relevant and engaging content, a well-articulated instructional model, effective teaching presence, learner support, and an enabling learning environment' (UNICEF, 2020). However, such use of technology requires teachers to be equipped with a whole new skill set and perspective connected to application of digital competencies in teaching and learning. In this paper, we refer to these skills and perspectives as a teacher's digital pedagogy competence and develop the Digital Pedagogy Competence Scale (DiPeCoS) to assess it.

## 2. What is Digital Pedagogy?

An analysis of literature (Kivunja, 2013; Montebello, 2017; Sailin & Mahmore, 2018) highlights that mere integration of technology in teaching and learning does not qualify as digital pedagogy. Rather, the purpose of technology integration must be to enrich or enhance learning. Such an intentional use of technology warrants technological skills, pedagogical skills as well as the ability to integrate both.

Given that the purpose of enhancement or enrichment of learning is at the center of digital pedagogy, it is essential to define what ‘enhanced’ or ‘enriched’ learning entails. Although defining and measuring the quality of teaching and learning is complicated (Fink, 2003; Berman, 2003; McCabe and Layne, 2012), it is widely agreed that learners have variable characteristics, preferences, needs, and abilities which need to be considered in the design of a learning experience to improve the experience for all (Al-Azawei et al., 2016).

Universal Design for Learning (UDL) is a framework to improve and optimize teaching and learning for all learners based on insights from cognitive neuroscience (CAST, 2015). Literature suggests a positive relationship between application of UDL and student interest and engagement (Smith, 2012) as well as its potential to improve students’ academic and social outcomes (Ok et al., 2017). There are 3 principles of UDL which guide the design of learning to suit a wide variety of learners. These are:

- Providing multiple means of engagement: This emphasizes the ‘why’ of learning and offers checkpoints on recruiting interest, sustaining effort & persistence and self-regulation.
- Providing multiple means of representation: This focuses on the ‘what’ of learning and captures checkpoints under - perception, language & symbols, and comprehension.
- Providing multiple means of action & expression: This focuses on the ‘how’ of learning and captures checkpoints under physical action, expression & communication, and executive function

Integrating digital tools in alignment with the UDL principles enables teachers to intentionally use technology to enhance learning. Such intentional use of digital tools in alignment with the three principles of UDL can reduce learning barriers and support students to meet learning and affective goals (Rao, 2021).

Using Universal Design for Learning as the reference framework, we propose that digital pedagogy entails leveraging digital technology with the purpose of: one, presenting information such that it can be perceived and comprehended by learners effectively; two, offering multiple strategies to engage learners such that they are motivated to learn; and three, enabling learners to navigate the learning environment and express what they know.

### 3. Existing Scales to Assess Digital Pedagogy Competence

Many tools have been developed to evaluate teacher’s digital competence. Beyond pedagogical skills, these tools include application of technology in a range of work teachers carry out in the classroom, the educational institution, the community and in the context of their own personal and professional development (Lázaro-Cantabrana et al., 2019). Literature review of existing tools like the Self-reflection on Effective Learning by Fostering the use of Innovative Educational technologies (SELFIE) tool based on DigCompEdu (Redecker & Punie, 2017), Teachers’ Digital Competencies Questionnaire based on the Common Digital Competence Framework for Teachers by INTEF (Tourón et al., 2018), Wayfind Teacher Assessment based on International Society for Technology in Education (ISTE) standards for teachers (Banister & Reinhart, 2013), COMDID based on other frameworks of teacher’s digital competence (Lázaro & Gisbert, 2015) points that most tools measure teacher’s digital competence, a concept that goes beyond digital pedagogy competence, and only a few tools have a targeted focus on evaluating a teacher's digital pedagogy competence.

Two tools that are closely related to the concept of teacher's digital pedagogy competence are the Survey of Preservice Teachers’ Knowledge of Teaching and Technology (SPTKTT) by Schmidt et al. (2009) which is based on the TPACK model (Mishra & Koehler, 2006) and the UDL self-assessment tool by University of Waikato (2018) based on the UDL framework (CAST, 2015). Both the tools use Likert scale-based items to get teachers to reflect on their pedagogical practices (for e.g., “*I know how to select effective teaching approaches to guide student thinking and learning in mathematics*” on the SPTKTT and “*I encourage students to express their learning in multiple ways (e.g., essay, or video blog, poster or presentation)*” on the UDL tool). The problem with these tools is that they rely on self-report of respondents’ behaviors, beliefs, perceptions, attitudes, or intentions, which are shown to be virtually uncorrelated with their on-the-job behavior (Thalheimer, 2018). Responding on self-report scales is also often affected by biases of social desirability (Van de Mortel, 2008), which can corrupt collected data. Thus, although self-report tools can be used for reflection and self-assessment that may

promote learning and improvements in performance (Andrade, 2019), they cannot adequately assess a teacher's digital pedagogy competence.

Any tool designed to measure competence must focus on the ability to apply knowledge and skills in real-life context. Thalheimer (2018) argues that assessing decision making is better than gauging self-perception of skills or behaviors. He also argues that one of the ways to evaluate realistic decision making is by presenting learners with realistic scenarios and prodding them to make decisions that are similar to the types of decisions they will have to make on the job. It is this form of scenario-based assessment that efficiently counters the shortcomings of self-report tools. Scenario-based assessments are based on the situated learning theory (Lave & Wenger, 1991) which states that learning and assessment best take place in the context in which they are going to be used (Kindley, 2002) and thus can be more efficient measures of assessment.

#### **4. Purpose**

There is a need for a tool to assess teacher's digital pedagogy competence which does not rely on self-report but rather makes use of scenario-based assessments in which respondents apply their subject knowledge, critical thinking and problem-solving skills in a real-world context to respond reliably. The Digital Pedagogy Competency Scale (DiPeCoS) for teachers has been developed to address this need. The following sections present the results of the validation of the DiPeCoS.

#### **5. Materials and Methods**

##### *5.1 Participants*

A total of 1315 English-speaking Indian teachers completed the virtually delivered scale. Participants reported a mean age of 42.1 years and had all received formal education in English language. 39% (N = 513) of the participants reported their gender as female and 61% (N = 802) as male. 3.6% of the teachers reported the setting of their work as "primary school", 28.5% as "middle school", 52.6% as "high school" and 15.3% as "others." Participants who reported their work setting as "others" included education counselors, consultants and freelance teachers. Teachers reported a mean teaching experience of 14.77 years (SD = 10 years, range 0-50 years).

##### *5.2 Item Development*

As mentioned above, DiPeCoS is based on the Universal Design for Learning (UDL) framework. It assesses the competence of teachers to purposefully use digital technologies to: 1) Present information such that it can be perceived and comprehended by learners effectively; 2) Offer multiple strategies to engage learners such that they are motivated to learn; and 3) Enable learners to navigate the learning environment and express what they know.

Ten multiple-choice, scenario-based items were developed to assess the ability of the respondents to leverage digital technology for the three pedagogical purposes mentioned above. For example, item 8 evaluated a teacher's ability to design a learning experience by using digital books and pointer tools to teach language was developed to assess the ability to use technology to represent information such that it can be perceived and comprehended by learners effectively. This purpose of technology integration is aligned with the UDL principle of multiple means of representation. Similarly, item 6 was developed to assess a teacher's ability to use technology to design a peer learning experience for 40 students to learn science at home. This item evaluated the use of technology to engage all the learners such that they are motivated to learn, this is aligned with the second principle of UDL, namely multiple means of engagement.

The pedagogical decision of the respondent in each scenario serves a dominant purpose and at the same time, enables other purposes too. For instance, while item 1 was developed to assess the ability of a teacher to leverage technology to facilitate peer sharing, which is aligned with the purpose of enabling learners to express what they know, peer sharing is also a strategy to engage and motivate learners and therefore serves a secondary function of engaging learners. Of the 10 items on the original scale, 3

questions were designed to assess the use of digital technology to represent information, 3 to engage learners and 4 to enable learners to navigate the learning environment and express what they know.

Popular guidelines on item development were followed, and it was ensured that the items were unbiased in terms of responding by government and private school teachers, as well as other forms of diversity. Items of the Digital Pedagogy Competency Scale (DiPeCoS) and their alignment to UDL principles can be found here: <https://bit.ly/3BQOxws>.

### 5.3 Procedure

Participants recruited for the validation of the tool were a part of a wider intervention that included taking an online course on digital pedagogies, and the scale developed in this study was utilized as a pre-post questionnaire to assess the impact of the course-based intervention. The results reported in this study were obtained from the pre-assessment responses. Recruited participants involved English-speaking teachers from English-medium schools across various parts of India, who filled out the questionnaire over a period of 4 months, from January to April 2022. Teachers working at private schools in urban areas as well as government-run residential schools operating in various parts across rural India were both a part of the study. Since many teachers did not provide details about their schools, it was not possible to determine the exact percentage of the different types of schools.

All participants were briefed about the aim of the study and given the opportunity to resolve queries related to the study and the questionnaire during a 1.5-hour long online workshop. After the workshop, participants created an account on Framerspace, an interactive learning platform ([www.framerspace.com](http://www.framerspace.com)), where the questionnaire and the course were hosted. The questionnaire consisted of the scale described above, along with a short demographic form that collected demographic information like gender, age, teaching experience and teaching profile. Information on time required to complete the assessment was provided. Informed consent was sought from the participants. No personally identifiable information (such as name or email address) was collected from the participants. Instead, entries were recorded, stored and identified using anonymized IDs provided to all participants, which could not be linked to their identities.

### 5.4 Data Analysis

A total of 1824 entries were received on the questionnaire. After removal of duplicates, 1315 entries remained, which were used to perform a validation of the scale. All statistical analyses were performed on R version 4.0.2 (<https://www.r-project.org>). An item response analysis was conducted to validate the developed scale because Item response theory (IRT) is useful to investigate if the items in a scale do not have enough reliable information about the construct being measured. It can also differentiate item properties (e.g., discrimination and difficulty) among individuals across a much wider range of the construct at hand. If the analyses show that there is such a problem with some items, the researcher can remove/modify those items or add new items that help to measure these parts of the construct, thus, providing information that can differentiate people across a much greater range of the latent trait and increase the validity of the whole scale (Oishi, 2007).

To use IRT, basic assumptions pertaining to unidimensionality, local independence, monotonicity and differential item functioning (DIF) were first tested. Unidimensionality (items in the scale load on only one latent factor) was tested using factor analysis. Exploratory factor analysis (EFA) procedures such as eigenvalue extraction (Kaiser, 1960), scree test (Cattell, 1966) and parallel analysis (Horn, 1965) were used to test the presence of a unidimensional factor, which was later affirmed using confirmatory factor analysis (CFA). Goodness of fit for the CFA model was evaluated based on commonly used indices:  $\chi^2_{S-B}/df < 5$  (Marsh & Hocevar, 1985), CFI, TLI  $> 0.90$  (Bentler, 1990), RMSEA  $< 0.08$  (Browne & Cudeck, 1992) and SRMR  $< 0.08$  (Browne & Cudeck, 1992). All factor analytic procedures used polychoric correlations with Satorra-Bentler correction (Satorra & Bentler, 1994) given the dichotomous nature and non-normality of items (Morata-Ramírez & Holgado-Tello, 2013). Local independence was tested by examining if the chance of one item being answered was related to any other item(s) being answered or if responses to items were independent decisions taken by the test-takers. Monotonicity (meaning that the levels of a person's latent trait increase, as a monotonic function, as the probability to choose the answer in each item that represents the participant's actual

level of the trait increases) was tested using Mokken analysis and DIF was applied to investigate whether the item responses varied across gender.

After assumption testing, an appropriate IRT model was selected. Since the items were scored on a dichotomous scale (0 for incorrect response and 1 for correct response), a logistic IRT model was chosen for the study. From the 1-PL, 2-PL and 3-PL models, a 3-PL model was chosen since 3-PL models, even though more complex, make it worthwhile sacrificing parsimony by providing the best fit. Moreover, the 3-PL model is considered appropriate for multiple-choice tests (like the one in this study) where the probability of success from a very low-ability person on an item may be significantly higher than zero because of random guessing (Diamond & Evans, 1973). Generally speaking, both the 2-PL and 3-PL models are considered more suitable for cognitive tests as compared to 1-PL models.

Using the 3-PL model, item parameters were calculated, and item information curves (IICs) and item characteristic curves (ICCs) were plotted for all the items. Results of these analyses, complemented with the theoretical judgment of the researchers, were used to guide removal of items that were problematic and increase the overall validity of the scale. After the scale structure was finalized, the test information function was plotted to observe how the overall scale responded to individuals with different abilities. Finally, a reliability analysis was performed. All results reported in this paper use a  $p$ -value of  $\alpha = 0.05$ .

### 5.5 Assumption Testing

We used EFA and CFA procedures to examine if the correlation among items in the scale was explained by a single latent factor. An unrestricted factor solution indicated that the magnitude of the first eigenvalue (4.11) was much greater than the magnitude of other eigenvalues (1.21, 0.91, 0.83, 0.72, 0.61, 0.56, 0.46, 0.36 and 0.18). The ratio of the first to the second eigenvalue was also greater than 3, which provided some evidence of unidimensionality (Sattelmayer et al., 2017). Complementing this, results from the scree plot and parallel analysis also hinted at single-factor solutions. A single-factor CFA model was thus built to confirm unidimensionality in the scale. The model converged normally and demonstrated a good fit:  $\chi^2_{S-B}/df = 0.667$ , CFI = 0.998, TLI = 0.998 and RMSEA (90% C.I.) = 0.000 (0.000 - 0.066). All item loadings were significant at  $p < .05$ , except items 7 and 9.

Local independence was confirmed as: a) the chance of one item being answered was not related to any other item(s) being answered, and (b) the response to an item was every test-taker's independent decision, i.e., there was no cheating or group work involved.

Results from Mokken analysis indicated that the response function of the probability of getting a correct response on each item increased when a person's latent trait increased (for all items except item 9). Therefore, evidence of monotonicity was seen in all items except item 9.

Results of a DIF analysis indicated that all items were responded similarly by both males and females. Next, we calculated the information, difficulty, discrimination and guessing parameters for all items in order to decide whether they needed to be excluded from the scale.

### 5.6 ITR Model

A 3-PL IRT model was created and discrimination, difficulty and guessing parameters were calculated for each of the 10 items. All items fitted well in the model ( $p > .05$ ). Discrimination parameter values can range from  $-\infty$  to  $+\infty$ , but values typically fall in the range of 0 to +2.50. Item discrimination values of 0.01–0.34 are considered very low; 0.34–0.64 low; 0.65–1.34 moderate; 1.35–1.69 high; and 1.70 and above very high (Baker, 2001). Item difficulty estimates vary from -4 to 4, where -4 represents most easy, 0 represents average and +4 represents most difficult. The guessing parameter,  $c$  has a theoretical range of [0,1], but in practice, values above 0.35 are not considered acceptable. As seen from Table 1, item 5 demonstrated “low” discrimination value, items 1, 2, 4 and 7 demonstrated “medium” discrimination values, item 8 and 10, “high” and items 3, 6 and 9, “very high.” Broadly speaking, items 1, 2, 3, 4, 6, 8 and 10 were “easy” items and items 5, 7 and 9 were “difficult” (based on the polarity of the difficulty parameter). Items 7 and 9 demonstrated high guessing parameters.

Table 1. 3-PL IRT model for the scale items

Item	Guessing	Difficulty	Discrimination	X <sup>2</sup>	P (> X <sup>2</sup> )
Item 1	0	-0.355	1.18	31.2758	0.8713
Item 2	0	-0.506	1.081	52.6484	0.1782
Item 3	0.001	-1.538	3.619	42.2773	0.5347
Item 4	0	-1.058	1.139	34.2256	0.5941
Item 5	0	0.967	0.565	25.2309	0.5842
Item 6	0	-1.063	2.266	45.2733	0.9406
Item 7	0.432	1.236	1.08	23.3789	0.5842
Item 8	0.002	-0.965	1.64	46.9475	0.703
Item 9	0.227	1.448	7.549	58.4472	0.4851
Item 10	0	-0.815	1.552	56.7459	0.4455

Item characteristic curves (ICC) and item information curves (IIC) were also plotted for the scale items (Figures 1 and 2, respectively). For both the graphs, theta ( $\theta$ ) represents a person's true latent trait (factor), which has been standardized to follow a normal distribution with a range from -4 to 4, where 0 represents the average score (Baker, 2001). For the ICC,  $P(\theta)$  represents the probability of a correct answer, while for the IIC,  $I(\theta)$  represent the information function, or how well each item contributes to score estimation precision (higher levels of information leading to more accurate score estimates). In an ICC, discrimination is defined as how well an item can differentiate between examinees having abilities below the item location and those having abilities above the item location. Consequently, items with ICCs which are more "spread out" indicate lower discriminability, ICCs which are farthest on the plot indicate higher difficulty, and ICCs that have a finite value of y-intercept indicate that there is a higher probability of guessing. In a similar way, IICs peak at the difficulty value (point where the item has the highest discrimination), with less information at ability levels farther from the difficulty estimate. As seen from Figure 1, items 5, 7 and 9 demonstrated higher levels of difficulty since they were placed on the right-hand side of the graph, indicating that the probability of responding to these items correctly would be high only for individuals with high ability. On the contrary, item 3 showed the lowest level of difficulty. Items with curves that were the least spread, for example, items 3, 6, 8, 9 and 10, indicated highest levels of discrimination. Since items 7 and 9 also had significant positive y-intercepts, they had a higher probability of being guessed. These results coincide with the ones inferred from Table 1.

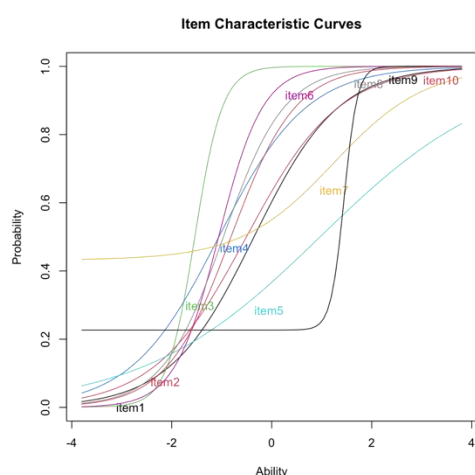


Figure 1. Item characteristic curves (ICCs) for the 10 items forming the scale

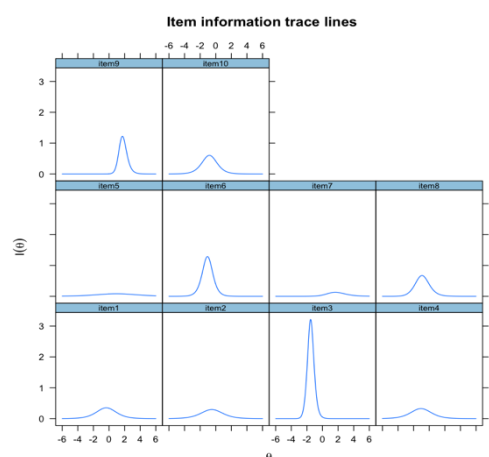


Figure 2. Item information curves (IICs) for the 10 items forming the scale

Some additional information was gathered from Figure 2. For instance, item 3 peaked very high at an ability level of  $\theta = -2.1$  and also demonstrated a narrow IIC, indicating that the item provided most information about low ability individuals. This was different as compared to, say item 4, which demonstrated a curve that was much wider spread and peaked at an ability level of  $\theta = -0.1$ .

A Test Information Function (TIF) was also plotted for the overall scale (see Figure 3a), which was simply the sum of information functions of all items of the scale. As seen in the figure, the TIF was

a bimodal curve, with two peaks at ability levels  $\theta = -1.8$  and  $\theta = 1.9$ . To the extent possible, the TIF should be a unimodal curve centred around  $\theta = 0$  so that the scale serves as an unbiased assessment of low and high ability individuals. In order to improve the quality of information provided by the scale, the researchers decided to remove items 3 and 9 based on the results obtained from the IRT analysis. Through a retrospective exercise, researchers concluded that it is possible that item 3 could be significantly affected by the bias of social desirability as response options (other than the correct one) provided on the item were “morally incorrect” for any teacher to answer (see details in Table 1). On the other hand, item 9 had ambiguity in the response options provided, due to which many of the responses could possibly be “correct.” After removal of these items, the TIF was re-plotted. As seen from Figure 3(b), the distribution improved. It demonstrated a unimodal peak around  $\theta = -0.2$  and was reasonably well spread. The 8-item scale was finalized.

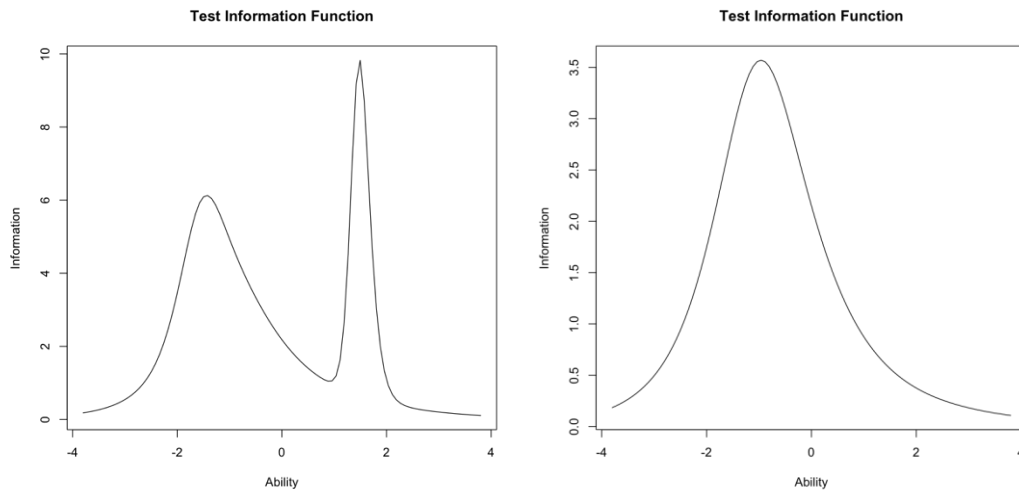


Figure 3. Test information curves (TIFs): a) before removing items 3 and 9, b) after removing items 3 and 9

### 5.6.1 Reliability analysis

As demonstrated in Figure 4, the scale demonstrated the highest reliability for individuals with ability levels around  $\theta = -1$ . There was almost no reliable information about below  $-2.5$  and about above  $2.00$ , and the standard error increased quickly for both smaller and larger  $\theta$  values. The marginal reliability for the scale was approximately 0.62.

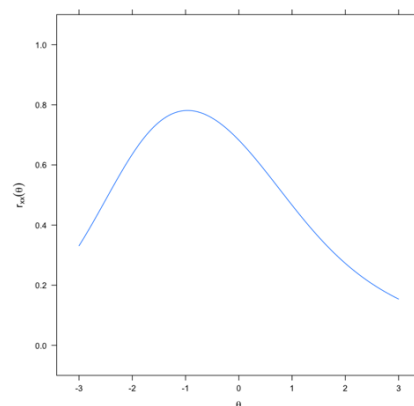


Figure 4. Reliability analysis of the scale

## 6. Discussion and Conclusion

Considering the demand for teacher’s digital pedagogy competence, especially after the rapid digitization in education due to COVID-19 lockdowns, it has become critical to define, measure and build teacher’s digital pedagogy competence. A literature review of the tools available to assess



teacher's digital pedagogy competence points at the need for a tool that does not rely on self-reporting to evaluate their digital pedagogy competence as data obtained from self-reports can be often unreliable. The Digital Pedagogy Competency Scale (DiPiCoS) has been developed to address this need and contains scenario-based items which are situated in real life learning and teaching contexts.

An item response analysis for the scale was conducted with responses from 1315 English-speaking teachers from various parts of India. The scale demonstrated unidimensionality and most of its constituent items showed acceptable levels of discrimination, difficulty and guessing parameters in an item response analysis constructed using a 3-PL model.

Results of the item response analysis showed that item 3, which was developed to assess whether or not a teacher practices inclusion in online classes, was answered correctly by most respondents. Through a retrospective exercise, researchers concluded that responses on this item were significantly affected by the bias of social desirability. On the other hand, item 9, which pertained to the integration of technology in pedagogical practice, had ambiguity in the response options provided, due to which many of them could possibly be "correct." Items 3 and 9 were removed from the 10-item scale and an 8-item scale was finalized. This scale demonstrated good reliability around ability levels  $\theta = -1$  and a marginal reliability value of approximately 0.62.

The validated DiPeCoS can be used as pre and post assessments for teacher training programs on digital pedagogy. The tool can also be used for screening purposes to identify educators with high and low digital pedagogy scores. With multiple teaching-learning scenarios being captured in the tool, the tool can be used with educators across K-12 to higher education institutions, across domains in any setting.

As future lines of research, external measures, which were not included in this study to save time and avoid respondent fatigue (Morgado et al., 2017), could be used to establish external discriminant and convergent validity of the scale. It might also be useful to examine some significant psychometric properties of the scale, such as test-retest reliability. Also, although assessment of responder's decision making with regard to pedagogical practices is a better measure of their digital pedagogical competence when compared with self-report, it still does not predict translation of these choices into real life practice.

Finally, the methods, strategies and goal of digital pedagogy continues to evolve with the emergence of new digital technologies and their affordances. Like any other tool designed to measure digital competence, this tool needs to be updated regularly to reflect new evidence and insights in the field of learning sciences that informs pedagogy as well as new advances in education technologies and their opportunities and constraints.

## References

- Andrade, H. L. (2019). A Critical Review of Research on Student Self-Assessment. In *Frontiers in Education* (Vol. 4). Frontiers Media SA. <https://doi.org/10.3389/feduc.2019.0008>
- Al-Azawei, A., Serenelli, F., & Lundqvist, K. (2016). Universal Design for Learning (UDL): A content analysis of peer reviewed journals from 2012 to 2015. *Journal of the Scholarship of Teaching and Learning*, 16(3), 39-56.
- Baker, F. B. (2001). *The Basics of Item Response Theory*. College Park: ERIC Clearinghouse on Assessment and Evaluation.
- Banister, S., & Reinhart, R.V. (2013). Assessing NETS•T Performance in Teacher Candidates: Exploring the Wayfind Teacher Assessment, 29(2), *Journal of Digital Learning in Teacher Education*.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238-246. <https://doi.org/10.1037/0033-2909.107.2.238>
- Berman, Elena. 2003. "A Short Guide to Evaluating Teaching." Tucson, AZ: University of Arizona. Accessed October 4, 2016. [http://www.tlataskforce.uconn.edu/docs/resources/A\\_Short\\_Guide\\_to\\_Evaluating\\_Teaching.pdf](http://www.tlataskforce.uconn.edu/docs/resources/A_Short_Guide_to_Evaluating_Teaching.pdf).
- Browne, M. W., & Cudeck, R. (1992). Alternative Ways of Assessing Model Fit. *Sociological Methods & Research*, 21(2), 230-258. <https://doi.org/10.1177/0049124192021002005>
- Cattell, R. B. (1966). The Scree Test For The Number Of Factors. *Multivariate Behavioral Research*, 1(2), 245-276. [https://doi.org/10.1207/s15327906mbr0102\\_10](https://doi.org/10.1207/s15327906mbr0102_10)
- Center for Applied Special Technology, CAST (2015), <https://www.cast.org/impact/universal-design-for-learning-udl>
- Diamond, J., & Evans, W. (1973). The Correction for Guessing. In *Review of Educational Research* (Vol. 43, Issue 2, pp. 181-191). American Educational Research Association (AERA). <https://doi.org/10.3102/00346543043002181>
- Fernández-Batanero, J.M.; Román-Graván, P.; Montenegro-Rueda, M.; López-Meneses, E.; Fernández-Cerero, J. (2021). Digital Teaching Competence in Higher Education: A Systematic Review. *Educ. Sci.*
- Fink, L. Dee. 2003. *Creating Significant Learning Experiences: An Integrated Approach to Designing College Courses*, Higher and Adult Education Series. San Francisco, CA: Jossey-Bass
- Horn, J.L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika* 30, 179-185. <https://doi.org/10.1007/BF02289447>



- Kaiser, H. F. (1960). The application of electronic computers to Factor Analysis. *Educational and Psychological Measurement*, 20(1), 141–151. <https://doi.org/10.1177/001316446002000116>
- Kindley, R.W.(2002). Scenario-based e-learning: A step beyond traditional e- learning. ASDT
- Kivunja, C. (2013). Embedding digital pedagogy in pre-service higher education to better prepare teachers for the digital generation. *International Journal of Higher Education*, 2(4), 131–142.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511815355>
- Lázaro-Cantabrana, J. L., Usart-Rodríguez, M., & Gisbert-Cervera, M. (2019). Assessing Teacher Digital Competence: the Construction of an Instrument for Measuring the Knowledge of Pre-Service Teachers. In *Journal of New Approaches in Educational Research* (Vol. 8, Issue 1, pp. 73–78). University of Alicante. <https://doi.org/10.7821/naer.2019.1.370>
- Marsh, H. W., & Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First- and higher order factor models and their invariance across groups. *Psychological Bulletin*, 97(3), 562–582. <https://doi.org/10.1037/0033-2909.97.3.562>
- McCabe, Kimberly A., and Leslie S. Layne. 2012. “The Role of Student Evaluations in Tenure and Promotion: Is Effective Teaching Really Being Measured?” *The Department Chair: A Resource for Academic Administrators* 22 (3):17–20
- Mishra, P., & Koehler, M. J. (2006). Technological Pedagogical Content Knowledge: A Framework for Teacher Knowledge. *Teachers College Record*, 108(6), 1017–1054. [http://onezoneheights.pbworks.com/f/MISHRA\\_PUNYA.pdf](http://onezoneheights.pbworks.com/f/MISHRA_PUNYA.pdf)
- Montebello, M. (2017). Digital pedagogies for teachers’ CPD [Conference presentation]. IADIS International Conference on Educational Technologies, Sydney, Australia.
- Morata-Ramírez, M. de, & Holgado-Tello, F. P. (2013). Construct validity of Likert Scales through confirmatory factor analysis: A simulation study comparing different methods of estimation based on Pearson and polychoric correlations. *International Journal of Social Science Studies*, 1(1). <https://doi.org/10.11114/ijsss.v1i1.27>
- Morgado, F. F. R., Meireles, J. F. F., Neves, C. M., Amaral, A. C. S., & Ferreira, M. E. C. (2017). Scale development: ten main limitations and recommendations to improve future research practices. *Psicologia: Reflexão e Crítica*, 30(1). Springer Science and Business Media LLC. <https://doi.org/10.1186/s41155-016-0057-1>
- Oishi, S. (2007). The application of structural equation modeling and item response theory to cross-cultural positive psychology research, in *Series in Positive Psychology*. Oxford Handbook of Methods In Positive Psychology, eds A. D. Ong and M. H. M. van Dulmen, (New York, NY: Oxford University Press), 126–138.
- 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(2), 116-138.
- Rao, K., Torres, C., & Smith, S. J. (2021). Digital tools and UDL-based instructional strategies to support students with disabilities online. *Journal of Special Education Technology*, 36(2), 105-112.
- Redecker, C., & Punie, Y. (2017). *European Framework for the Digital Competence of Educators: DigCompEdu*. Luxembourg: Publications Office of the European Union.
- Sailer, M., Murböck, J., & Fischer, F. (2021). Digital learning in schools: What does it take beyond digital technology? In *Teaching and Teacher Education* (Vol. 103, p. 103346). Elsevier BV. <https://doi.org/10.1016/j.tate.2021.103346>
- Sailin, S.; Mahmor, N. (2018), Improving Student Teachers' Digital Pedagogy through Meaningful Learning Activities. *Malaysian Journal of Learning and Instruction*, v15 n2 p143-173 Dec 2018
- Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye, & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399-419). Thousand Oaks, CA: Sage
- Sattelmayer, M., Hilfiker, R., Luomajoki, H., and Elsig, S. (2017). Use of Rasch analysis to investigate structural validity of a set of movement control tests for the neck. *Musculoskelet. Sci. Pract.* 27, 131–136. doi: 10.1016/j.math.2016.07.006
- Schmidt, D. A., Baran, E., Thompson A. D., Koehler, M. J., Mishra, P. & Shin, T. (2009). Technological pedagogical content knowledge (tpack): The development and validation of an assessment instrument for preservice teachers. *Journal of Research on Technology in Education*, 42(2), 123-149. doi: 10.1080/15391523.2009.10782544
- Smith, F. G. (2012). Analyzing a college course that adheres to the Universal Design for Learning (UDL) framework. *Journal of the Scholarship of Teaching and Learning*, 31-61.
- Thalheimer, W. (2018). The learning-transfer evaluation model: Sending messages to enable learning effectiveness. Available at <https://WorkLearning.com/Catalog> This report is aligned with LTEM Version 11.
- Tourón, J., Martín, D., Navarro Asencio, E., Pradas, S., & Íñigo, V. (2018). Validación De Constructo De Un Instrumento Para Medir La Competencia Digital Docente De Los Profesores (Cdd). In *Revista Española De Pedagogía* (Vol. 76, Issue 269). Funciva Ediciones Sl. <https://doi.org/10.22550/Rep76-1-2018-02>
- UNICEF. (2020). Putting the ‘learning’ back in remote learning. <https://www.unicef.org/globalinsight/reports/putting-learning-back-remote-learning>
- University of Waikato. (2018). Universal Design for Learning Self-Assessment. [https://www.waikato.ac.nz/\\_data/assets/pdf\\_file/0013/501007/for-web-UOW-UDL-self-assessment-tool-draft-printable-v3.1.pdf](https://www.waikato.ac.nz/_data/assets/pdf_file/0013/501007/for-web-UOW-UDL-self-assessment-tool-draft-printable-v3.1.pdf)
- van de Mortel, TF 2008, 'Faking it: social desirability response bias in self-report research', *Australian Journal of Advanced Nursing*, vol. 25, no. 4, pp. 40-48.