Journal of Inclusive Methodology and Technology in Learning and Teaching
ISSN 2785-5104
Anno4 n.3(2024)

Edizioni Universitarie Romane
editoria scientifica dal 1952

A digital gamified approach to assess problem-solving skills in K9

Arianna Marras[1]* and **Masiar Babazadeh**[2]

[1] Department of Humanities, Languages and Cultural Heritage, University of Cagliari, Cagliari, Italy; arianna.marras@unica.it
[2] Department of Innovative Technologies, University of Applied Sciences and Arts of Southern Switzerland, Lugano, Switzerland; masiar.babazadeh@supsi.ch
* Correspondence: arianna.marras@unica.it

Abstract: Interest in measuring students' executive functions has significantly increased over the past two decades due to their importance as predictors of academic success. This study tests an application, the Digital Tower of London (ToL), for measuring problem-solving and planning skills in school. The results of this study show that the digital ToL does not reveal statistically significant differences in accuracy scores assessing problem-solving ability compared to the tangible test, and that students (11-12 years old) appreciate the assessment in the form of a game on a tablet. However, despite the limited sample of 57 students, the potential of new digital tools for measuring executive functions is evident. These tools are self-assessed and sustainable within the school environment as they offer teachers the opportunity to integrate information on cognitive abilities into students' learning paths.

Keywords: Problem solving Assessment; Digital Tower of London; Educational Technology; Executive Functions; Inclusive education.

1. Development, enhancement and assessment of executive functions at school

Historically, video games are designed without any educational intent, as they already have significantly transformed this field. Educational games can help leverage the teach- ing and training of various subjects, capitalizing on their potential engagement power. In fact, concepts as gamification (the use of game elements in non-gaming contexts) and game-based learning (learn by playing) are used as motivational tools to improve learning. (Video) Games have the possibility to give immediate feedback, allowing students to immediately a knowledge if they gave a correct answer as a result of their thought process (Gee, 2005), offering a formative evaluation tool and a secure learning environment where errors do not have a real negative impact (Squire, 2005), stimulating resilience (McGonigal, 2011).

Other than learning per se, video games can be used as a tool to assess students competencies in topics, and/or soft skills.

By designing games that require students/players to use their knowledge in order to achieve an in-game goal, they can be leveraged to evaluate subject matter competencies through questions/quizzes within the game, or actions to be performed in a game in a given order. Authentic assessment can be achieved by designing games as simulations of close-to-real-life situations, in which actions performed in the game are closely related to real life actions (Barab & Dede, 2007), be it in Virtual Reality or not.

On the one hand, interest in video game approaches as educational devices in teaching is continuously growing in terms of enhancement and assessment of both disciplinary learning (Giannandrea & Winstone, 2024) and specific cognitive functions. On the other hand, it is well known that good design of multimedia tasks stimulates the activation of cognitive processes and supports learning processes (Mayer & Moreno, 2016).

Several studies have investigated the impact of executive functions (EF) on developmental progress as predictive factors for academic success (Diamond, 2013). Executive Functions (EF) can essentially be considered a set of interconnected abilities that allow each person to complete goal-directed and adaptive actions (Lezak, 1983; Gioia *et al.*, 2001), by controlling cognition and regulating behavior. We refer to Adele Diamond's (2013) EF model, which identifies three basic components: shifting, updating, and inhibition, which are interrelated and interdependent. Although EF are presented and discussed as distinct functions, literature widely supports the idea that they work integratively in daily life (Diamond, 2016) and, therefore, in school life, allowing students to effectively engage in complex processes, such as defining a task-solving strategy. This highlights the profound involvement of EF in aspects related to learning and academic performance (Zelazo *et al.*, 2016).

Limited development of EF can lead to learning difficulties that may also result in behavioral issues (Vicari & Di Vara, 2017). Given the significant influence of EF on academic success, it is useful to assess their development in formal education, thus we face the need to develop EF measurement tools suitable for primary school students that are engaging, accessi ble, and consider appropriate completion times and alternative administration methods. Computerized tasks and tablet applications allow students to work independently by following verbal and visual instru ctions on supports such as touch screen. Considering the advantages of using tablets over computers (Falloon, 2013), including reliability and speed of measurement (Howard & Okely, 2015), and the limited research on this topic, it is crucial to u nderstand the students' perceptions using these applications to assess their EF. Berg *et al.* (2020) reported that children enjoyed performing activities with EF assessment apps, and children who enjoy tasks perform better (Schukajlow & Krug, 2014). However, we are currently facing several challenges. It is necessary that game-like applications are reliable and follow a rigorous study of psychometric properties. When facing a problem to solve, we activate a series of cognitive processes: pla n- ning a useful strategy, using working memory to access information, and implement the planned solution. If the planned solution does not work, we use cognitive flexibility to make adjustments along the way, and not to repeat the previous incorrect response.

To bridge the existing gap in the literature, it is not only necessary to design new tasks for measuring EF but also to capture the perceptions of children and adolescents regarding the tests and evaluate the students' experience (Berg *et al.*, 2020).

The present work involves the experimentation and verification of a digital application for evaluating problem-solving skills. The development of a mobile application is an integral part of a broader project, born through an international collaboration, and focuses on the implementation and verification of the potential of an edugame as an educational tool for the development and assessment of cognitive skills in schools.

2. The current study, evaluating a Digital Tower of London: a problem-solving measurement tool

This work aims to improve a digital game inspired by the tasks of the Tower of London (ToL) test, and therefore capable of assessing problem-solving and planning skills of students in the first cycle of school. The ToL test is a widely used assessment in applied clinical neuropsychology and it is used to evaluate executive functions, particularly identifying deficits in planning. The Digital ToL would not only qualify as an innovative tool in the field of compulsory formal education but could also stimulate teacher engagement in learning and experimenting with new technology-mediated approaches. Like the validated tangible version of the test (Fancello *et al.,* 2013), the Digital ToL has been developed to assess problem-solving and planning abilities.

One of the main differences compared to the Tangible ToL concerns the administration mode. It is not necessary to have an examiner alongside the student, thus the ratio examiner-examinee is not 1:1 because the system can autonomously manage data collection (correct answer scores, times, number of attempts, and violations). Thus, teachers can administer the Digital ToL and evaluate problem-solving skills of their students just by launching the application on the students' digital devices.

Along with studies emerging in the literature (e.g., Falloon, 2013), the aim is to show the task's objective, providing the possibility of a preliminary exemplified trial (without involving, for the time being, actions such as hand movements or verbal explanations) that serves as a practical example before the test begins. All these aspects, including the interaction parameters imposed on students (such as the inability to pause), the possibility of incorporating text-to-speech to include verbal and non-verbal representations (Paivio, 1991), are part of the ongoing study on the app's design features, which is still a work in progress due to its iterative and co-design nature with educational stakeholders.

This research, in line with the concerns of Maddux and Cummings (2004) regarding the lack of theory guiding the adoption of new technologies in schools, embraces the new theory on the design of software for student learning in combination with pedagogical-didactic aspects (Falloon, 2013). In this sense, as found in our previous work (Marras & Babazadeh, 2023), it is a collaborative design that has provided and continues to offer the possibility to meet and share ideas among teachers, developers, students, and researchers to improve the learning experience mediated by the technological device.

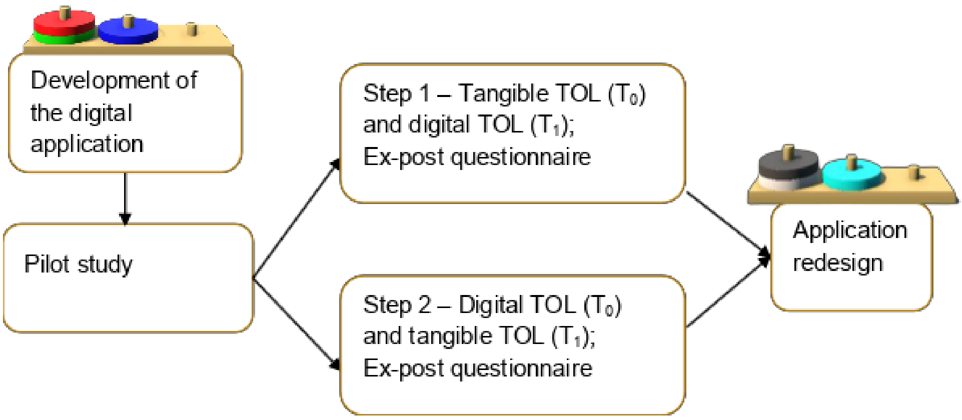
3. Materials and Methods

This research, inspired by Design-Based Implementation Research (Fishman *et al.*, 2013), has given rise to an iterative cycle of app design, testing, evaluation through the analysis of student performance, and debriefing with participants and their teachers.

This work has a dual purpose: firstly, to investigate whether there are statistically significant differences in the accuracy of task performance in the ToL test depending on the modality in which the task is presented; secondly, it seeks to understand the perception of the students involved in the study regarding the two different approaches. The results of the first phase of our study (Marras & Babazadeh, 2023) highlighted the ability of the Digital ToL to evaluate the constructs investigated. Our hypothesis aims to verify whether there are significant differences by reversing the test execution and asking students to solve the problems first with the Digital ToL and subsequently with the Tangible (Figure 1).

The Digital ToL may be more engaging for students due to its playful nature and may make students feel more comfortable and motivated in solving tasks. The aim is to verify aspects related to students' perceptions of the two modalities of presentation. The use of the Digital ToL could be more sustainable in the school context, because it can be presented to the entire class at the same time, and because it helps to limit the impact of other variables such as the subjectivity of the administrator and errors in data collection and scoring.

Figure 1. *Research phases.*



3.1. Participants

We examined two different groups of students, each one presented with a double administration of the ToL. The first phase is detailed in Marras & Babazadeh (2023). The second phase involved another group of students (N=31) who were asked to solve the tasks first with the Digital ToL and then with the Tangible ToL after the same amount of time.

Table 1. Master information of research sample

Sample Gender	Pilot Study				Tot N
	Step 1		Step 2		
	N (%)	**Mean age (years)**	**N (%)**	**Mean age (years)**	
F	15 (57,7%)	11.85	8 (25.8%)	12.1	23
M	11 (42,3%)	11.65	23 (74.2%)	12	34
Tot.	26 (100%)	11.77	31 (100%)	12	57

3.2. Research procedure

Before starting, we requested approval from the participating school. Information and consent forms were sent to the school principal, the teachers involved in the study, and the parents. Appropriate times and dates for data collection were established, and conducted in a school environment familiar to the students (the computer lab or the classroom). A total of 57 students from a middle school in South Sardinia were recruited for this study. The students were divided into two groups; one was presented the Tangible ToL and, after a month and a half, the Digital ToL, and the second group first the Digital ToL and then the Tangible ToL. After working in both modalities, students were asked to complete a questionnaire investigating their perception of the experience.

3.3. Survey tools and administration

The Tower of London (Norman & Shallice, 1986; Shallice, 1982) evaluates problem-solving ability in both tangible and digital formats. The tangible variant (Fancello *et al*., 2013) is administered individually in about 15 minutes per student. Students are invited to solve tasks using a tool with three rods of decreasing length that can hold respectively three, two, and one sphere, each of a different color. The test consists of twelve different problems of increasing difficulty, starting from an initial situation where the spheres are positioned on the three rods, the number of moves allowed to solve the task, the target situation to be achieved, and the maximum number of attempts possible for each problem. During the test the administrator records the time taken for "decision-making" (the time between the presentation of the problem and when the student picks up the first sphere to make a move) and "execution" (the time from the first move to the last, which may coincide with the solution to the problem or the last allowed move). If the rules, such as no two spheres can be handled simultaneously, are violated, the student is asked to reposition the spheres in the initial position and proceed to the next attempt. Solving the problem on the first attempt earns the student a score of three points, on the second attempt a score of two points, and one point on the third attempt.

3.4. Data Analysis

The data were analyzed using the Jamovi software for MacOSX (release 2.3.21). Differences among the administration of the Tangible and Digital versions were explored through a paired t-test. A descriptive statistical analysis was conducted on the students' perceptions.

4. Results

4.1. Accuracy in ToL task execution: Tangible vs Digital

During the data collection phase we gathered useful information about the scores obtained during the ToL in terms of accuracy and timing. Inferential analyses focused on exploring the differences between the mean scores obtained in ToL problems in the two different experimental phases. The data showed no significant differences between the averages. Further analysis with the paired t-test (Table 2) showed no statistically significant differences in the accuracy scores obtained in both tangible and digital tasks in step 2 (t(30)=-1.26, p=.217), confirming the data collected previously in step 1 (t(25)=0.0415, p=.967).

Table 2. Paired sample t-test in ToL: problem solving accuracy, decision, executive and total time.

Factors	ToL	Pilot study: STEP 2 ($DigT_0\_TangT_1$)				
		Mean	SD	t	df	p
Problem solving accuracy	Tangible	28.2	3.69	-1.26	30	0.217
	Digital	27.3	3.72			
Decision Time	Tangible	67.6	27.98	-2.65	30	0.013
	Digital	51.8	15.59			
Executive Time	Tangible	98.2	34.57	3.91	30	< .001
	Digital	147.6	57.35			
Total Time	Tangible	165.8	38.38	2.77	30	0.010
	Digital	199.4	61.43			

4.2. Feedback questionnaire Students' perceptions about experience: Tangibile VS Digital

Regarding the students' perception of the two methods, Table 3 suggests that there is less discouragement among students working with the digital tool.

Table 3. Student's perceptions of Digital and Tangible TOL

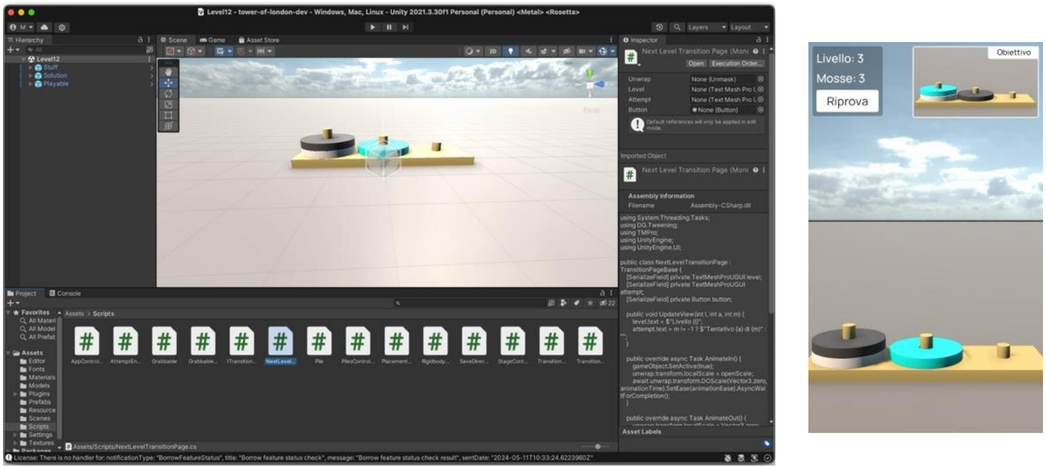
Aspects investigated	Pilot study							
	Step 1				Step 2			
	Digital		Tangible		Digital		Tangible	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Pleasure	4.19	0.69	3.96	0.96	4.29	0.74	4.19	0.83
Comfort	4.54	0.81	3.88	1.11	3.84	0.90	3.65	0.91
Enjoyment	4.23	0.71	-	-	4.16	0.90	4.32	0.83
Discouragement	-	-	1.88	0.99	1.94	1.00	2.13	1.09

The collected data (Table 3) suggests that the Digital ToL is generally preferred. However, when students were explicitly asked about their preference, 45.16% stated that they preferred the digital version, while the remaining 54.84% favored the Tangible mode. After the sessions, a higher number of students mentioned they had a better time and did not feel negative emotions after with the Digital ToL. Among those who preferred the Tangible ToL, some noted difficulties in performing the digital test due to technical issues. These consider ations led to the need for a r edesign of the application.

4.3. Application Re-design

The Digital ToL was redesigned and improved in Unity LTS 2022 (Figure 2). Positive reinforcements were added at the end of each task as immediate feedback, a short break was introduced between problems (Figure 3), and visual aspects such as the color of the disks were modified for accessibility. The redesign did not change the nature of the tasks, which remained the same, but focused solely on interface elements to make it more ergonomic and enhance its usability for students.

Figure 2. *Unity interface during redesign (a), redesigned application (b).*



(a) (b)

Some issues related to the movement of the disks, which occasionally conflicted with the game's physics, have been corrected to ensure a smooth experience for the students using the application.

Figure 3. *Breaks in between stages (a) and new feedback screens (b, c, d) were added to the redesigned application.*

(a) (b) (c) (d)

5. Discussion

Considering the limitations of existing approaches for executive development in the school context, this work attempts to preliminarily validate a digital application to measure problem solving skills, aiming to ensure an intuitive and engaging interface that reduces the influence of prior learning (e.g., literacy) on students' performance.

From the analysis of the students' performance data collected from the Tangible and Digital ToL, no statistically significant differences emerged between the two approaches. Therefore, it can be inferred that both measurement tools were capable of assessing problem-solving skills and that the digital ToL could be qualified as a teaching device with good potential, especially if implemented in the design of an educational game. Regarding the task completion times, the digital version showed a higher average in both execution time and total time. The second administration (Step 2) shows slightly lower averages compared to the first. The longer times recorded with the digital ToL suggested redesigning the application to make it more ergonomic and stimulating (Section 4.3). Students, while performing the digital task on the tablet without an educational figure to support them, were completely autonomous in managing their work time: this could justify the results recorded in execution time. Further administration of the application is deemed necessary following the redesign to verify if the aforementioned significant differences persist. A more detailed analysis of the times is also considered desirable, taking into account the problem-solving process planned by the students and integrating additional digital tools for measuring executive functioning (e.g., Berg et al., 2020) to evaluate whether the result times can be attributed to other cognitive processes.

It is interesting to note that the decision times for the Digital ToL are lower compared to the Tangible ToL (Table 2) in both steps of the pilot study. This is believed to be due to the attractiveness of using this tool in the classroom, social desirability, and the natural attitude of younger generations towards the use of digital devices. The short time between tasks (or attempts) may also have influenced this result. To determine if this finding is confirmed, further experimentation with the new application is necessary.

Students reported finding the digital tasks enjoyable, a finding supported by literature, which highlights a preference for digital assessments that do not involve the teacher as a test administrator, generally considered less stressful (Berg et al., 2020; Singleton et al., 2009). However, in the present study, students did not find the digital application more fun than the tangible one, although they reported feeling less discouraged when solving the digital task.

A digital measurement tool effectively meets the needs of schools and teachers, as it allows for large-scale administration and background data collection, providing teachers with important information on the executive functioning of their students. In this direction, the tool aims to meet educational needs by offering teachers an output that includes a profile of the student's functioning and the performance ranges of the class, which are useful in planning educational design.

6. Conclusions, limitations, and future work

The collected and analyzed data seem to support the hypothesis that the functions investigated by the ToL were adequately measured in both presentation modalities, despite the small sample of this exploratory study. This result encourages further verification of the digital tool's assessment capability using a larger sample.

Further possible modifications to the application are being reconsidered, such as the use of fonts to improve accessibility, and the saving of the tap and drag-and-drop positions to reconstruct the solving process.

In addition to planning a two-phase experiment with a sample of students of the same age and educational level to beta-test the redesign of the digital ToL, we plan a large-scale experiment to evaluate its psychometric properties. It could also be desirable to develop additional functions in the digital tool. Some, like the Digital ToL, would allow for the assessment of executive functions and problem-solving, while others would enable their training.

In conclusion, from the perspective of digital and inclusive education, the implementation of the Digital ToL could provide educational professionals with a valuable tool to support classroom teaching, addressing the gap identified in the literature on the limited use of digital tools in teaching practices (Zappalà & Viola, 2023). The tool, focusing on verifying aspects related to the development of students' executive functions, also offers teachers immediate feedback on students' abilities, giving teachers the opportunity to guide and redefine educational paths.

Acknowledgements
We extend our gratitude to the head of the South Sardinian school, as well as the teachers and students, for their enthusiastic participation and willingness to contribute to this research.

References

Barab, S. A., & Dede, C. (2007). Games and immersive participatory simulations for science education: An emerging type of curricula. Journal of Science Education and Technology, 16(1), 1-3.

Berg V., Rogers S.L., McMahon M., Garrett M. and Manley D (2020). A Novel Approach to Measure Executive Functions in Students: An Evaluation of Two Child-Friendly Apps. Front. Psychol. 11:1702. doi: 10.3389/fpsyg.2020.01702

Diamond, A. (2013). Executive functions. Annual Review of Psychology, 64(1), 135–168.

Diamond, A. (2016). Why improving and assessing executive functions early in life is critical. In Executive function in preschool-age children: Integrating measurement, neurodevelopment, and translational research. (pp. 11-43).

Falloon, G. (2013). Young students using iPads: app design and content influences on their learning pathways. Computers & Education. 68, 505–521. doi: 10.1016/j.compedu.2013.06.006

Fancello G.S., Vio C., Cianchetti C. (2013). TOL. Torre di Londra. Test di valutazione delle funzioni esecutive (pianificazione, problem solving). Erickson.

Fishman, B. J., Penuel, W. R., Allen, A.R., Cheng, B. H., & Sabelli, N. (2013). Design-based implementation research: An emerging model for transforming the relationship of research and practice. In Fishman, B. J., & Penuel, W. R. (Eds.), National Society for the Study of Education (pp. 136-156). New York, NY: Teachers College Record.

Gee, J. P. (2005). Learning by design: Games as learning machines. Interactive Educational Multimedia, (8), 15-23

Giannandrea, L., & Winstone, N. (2024). Valutazione e tecnologie in Tecnologie per l'educazione 2/ed.; Milano, Pearson; pp. 177 - 188 (ISBN: 9788891932150,9788891932167)

Gioia, G., Isquith, P., Guy, S., and Kenworthy, L. (2001). Brief Rating Inventory of Executive Function (BRIEF). Odessa, FL: Psychological Assessment Resources.

Howard, S. J., & Okely, A. D. (2015). Catching Fish and Avoiding Sharks: Investigating Factors That Influence Developmentally Appropriate Measurement of Preschoolers' Inhibitory Control. J Psychoeduc Assess, 33(6), 585-596. doi:10.1177/0734282914562933
Lezak, M. (1983). Neuropsychological assessment. Oxford University Press.lez
Maddux, C., & Cummings, R. (2004). Fad, fashion, and the weak role of theory and research in information technology in education. Journal of Technology and Teacher Education, 12(4), 511-533.
Marras, A., & Babazadeh, M. (2023). Un approccio videoludico alla valutazione delle funzioni esecutive in ambito scolastico: una "Torre di Londra" digitale. Journal of Inclusive Methodology and Technology in Learning and Teaching, 3, 1–14.
Mayer, R. & Moreno, R. (2016). A Cognitive Theory of Multimedia Learning: Implications for Design Principles.
McGonigal, J. (2011). Reality is broken: Why games make us better and how they can change the world. Penguin.
Norman, D. A., & Shallice, T. (1986). Attention to Action: Willed and Automatic Control of Behaviour. In R. J. Davidson, G. E. Schwartz, & D. E. Shapiro (Eds.), Consciousness and self-regulation (pp. 1–14). Plenum Press. http://wexler.free.fr/library/files/norman () attention to action. willed and automatic control of behavior.pdf
Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology / Revue canadienne de psychologie, 45*(3), 255–287.
Schukajlow, S., & Krug, A. (2014). "Are interest and enjoyment important for students' performance," in Proceedings of the Joint Meeting of PME 38 and PMENA 36, Vol. 5, eds C. Nicol, S. Oesterle, P. Liljedahl, & D. Allan (Vancouver: PME).
Shallice, T. (1982). Specific impairments of planning. Philosophical Transactions of the Royal Society of London, Part. B, 298, 199-209.
Singleton, C., Horne, J., & Simmons, F. (2009). Computerised screening for dyslexia in adults. Journal of Research in Reading, 32(1), 137-152.
Squire, K. (2005). Changing the game: What happens when video games enter the classroom?. Innovate: Journal of Online Education, 1(6).
Vicari, S., & Di Vara, S. (2017). Funzioni esecutive e disturbi dello sviluppo. Erickson.
Zappalà, E., & Viola, I. (2023). La prospettiva dei futuri docenti di sostegno sul potenziale didattico e inclusivo delle TIC e dei videogames. Journal of Inclusive Methodology and Technology in Learning and Teaching, 3(3). https://www.inclusiveteaching.it/index.php/inclusiveteaching/article/view/104
Zelazo, P. D., Blair, C. B., and Willoughby, M. T. (2016). Executive Function: Implications for Education. NCER 2017-2000. Washington, DC: U.S. Department of Education.