Papeles del Psicólogo / Psychologist Papers, 2022 Vol. 43(1), pp. 36-47
https://doi.org/10.23923/pap.psicol.2986
http://www.papelesdelpsicologo.es
http://www.psychologistpapers.com

The Use of New Technologies in Educational Assessments: Reading in a digital world

Javier Suárez-Álvarez[1], Rubén Fernández-Alonso[2], Francisco J. García-Crespo[3] and José Muñiz[4]

[1]Universidad de Massachusetts Amherst, Estados Unidos. [2]Universidad de Oviedo, España. [3]Universidad Complutense de Madrid, España. [4]Universidad Nebrija, España

Ser un lector competente en un mundo digital requiere una base sólida de Lectura, pero también la capacidad para pensar críticamente; una tarea pendiente para muchos estudiantes españoles. Las pruebas adaptativas informatizadas y los datos de proceso (información sobre las acciones que realizan los estudiantes al responder a la prueba) son especialmente importantes cuando se evalúan competencias como la Lectura. El objetivo de este trabajo es analizar cómo el uso de la tecnología está cambiando el concepto de Lectura y las formas para evaluarlo. Esto tiene implicaciones tanto para el alumnado español como para cualquier profesional encargado de interpretar y diseñar evaluaciones educativas. El investigador debe asegurar que el uso de los datos y la tecnología sea el adecuado para los objetivos de la evaluación y sirva de forma fiable, válida y justa a las personas involucradas, pero también del usuario saber cuándo, cómo y para qué utilizar los datos.

Palabras claves: *Tecnología, Datos de proceso, Tests adaptativos informatizados, PISA, Competencia lectora.*

Being a proficient reader in a digital world requires a strong reading foundation, but also the ability to think critically, which is a challenge for many students in Spain. Computerized adaptive tests and process data (information about students' actions when responding to the test) are especially important when assessing skills such as reading. This work aims to analyze how the use of technology is changing the concept of reading and the ways to evaluate it. This has implications for Spanish students and any professional in charge of interpreting and designing educational assessments. The researcher must ensure that the use of data and technology is adequate for the purposes of the assessment and that it works in a reliable, valid, and fair way for the people involved, but also the user must know when, how, and for what purposes to use the data.

Key words: *Technology, Process data, Computerized adaptive tests, PISA, Reading.*

For a long time, it has been thought that from millennials onwards, digital skills should be taken for granted. This is beginning to be questioned. There is no doubt that today's students use technology more, but do they know how to use it well? If the data show anything, it is that having more exposure to technology does not necessarily imply greater digital competence, and the differences are huge between students from advantaged and disadvantaged backgrounds (OECD, 2021a). Numerous studies have shown that younger generations may be more familiar with technology than previous generations; however, "digital natives" are not always necessarily equipped with the right skills in terms of accessing and using digital information (Breakstone et al., 2018; Macedo-Rouet et al., 2020; McGrew et al., 2018; OECD, 2011). Students need to use information and communication technologies (ICT) to access texts through search engines, use links and tabs, process information from multiple sources, evaluate the quality of information sources, detect potential conflicts, and resolve them. Definitions of what it means to be a competent reader are continuously updated to respond to changes in technology, society, economy, and culture, and with them the ways of assessing these competencies. The aim of this paper is to analyze the way in which the use of new technologies is changing the concept of reading and, consequently, the ways of assessing it. In this article, we address some of the most important implications of these changes for Spanish students, as well as for any professional in charge of interpreting or designing educational assessments that take advantage of the advances in technology. For more information on the impact of new technologies on other areas of assessment such as psychology and ambulatory assessment see Andrés et al. (2022), Fonseca-Pedrero et al. (2022), and Santamaria and Sanchez-Sanchez (2022) in this monograph.

Received: 1 November 2021 - Accepted: 21 December 2021
Correspondence: Javier Suárez-Álvarez, Ph.D.. *Senior Research Fellow, Center for Educational Assessment. University of Massachusetts Amherst. College of Education. 813 North Pleasant Street. Amherst, MA 01003. USA. E-mail: suarezj@umass.edu*

READING IN A DIGITAL WORLD: THE EVER-CHANGING CONSTRUCT

According to the latest data from the Program for International Student Assessment (PISA, 2018), the average 15-year-old student in Spain spends about 35 hours per week connected to the Internet—which represents an increase of 66% in 6 years, and these are data prior to the Covid-19 pandemic so this figure has possibly not reached its peak (OECD, 2021a). However, in the PISA tests, about one in three students in Spain responded that clicking on a link in a phishing email was appropriate or very appropriate[1]. Less than half of 15-year-olds in Spain are able to distinguish fact from opinion in the

[1] *PISA 2018 included a task inviting students to click on an email link from a well-known mobile operator and fill out a form with their details to win a smartphone, also known as a "phishing" email.*

PISA tasks[2]. Spain is below the OECD average on both indicators, although the difference is less than 10 percentage points (INEE, 2021b). The good news is that working on these skills in the classroom (for example, learning to detect biased information) is related to better reading proficiency (for example, being able to distinguish fact from opinion) (Suarez-Alvarez, 2021). These data confirm that the school can play a key role in learning digital skills and helping to reduce the risks associated with them, although, of course, there are no magic formulas.

While it is true that 29 of the 35 hours per week that Spanish students spend on the Internet are outside of school, the countries where students spend the most hours using technology in the classroom are not always the ones that show the best digital competence (OECD, 2021a). The relationship between reading achievement and time spent using digital devices for schoolwork was negative in 36 countries and economies, including Spain. In Australia, Denmark, Korea, New Zealand, and the United States, this relationship was positive after taking into account the socioeconomic situation of students and schools. Similarly, we must not forget that the use of digital technologies responds to a multidimensional reality: sometimes the use of technology in the classroom is complementary to other activities that are also important, while in other cases it can replace them. The key may lie not so much in the use itself, but in how a particular activity is carried out.

For example, there is still a clear benefit to reading on paper, which is something the literature that has investigated this issue agrees on (Clinton, 2019; Delgado et al., 2018). These results are also consistent across the more than 70 countries and economies that participated in PISA 2018: students who read books more frequently in paper than in digital format score better on the PISA 2018 reading tests and spend more time reading for pleasure. These data suggest that we should not banish paper in favor of a digital monopoly, and it seems reasonable to seek compatibility between traditional and analog reading formats. In fact, the most proficient readers are those students who are able to optimize the advantages of both formats. Students with good reading skills seem to respond to the following profile: they read books on paper and use digital devices, depending on the objective, i.e., they can read a novel on paper, but read daily news online.

Another clear example is the type of reading strategies that students learn in school, a fact that in the case of Spain is particularly interesting. Students in Spain scored particularly high on two of the three reading strategy knowledge indices included in PISA 2018 (comprehending and recalling a text, and writing a summary), but less so on the index of reading strategies for assessing the credibility of information sources (INEE, 2021). The data suggest that students in Spain have a relatively good knowledge of the traditional and still important aspects of reading. However, they still lack, on average, the relevant knowledge and skills to navigate in a digital world. Reading is not just about learning to decode written words during childhood, but about learning a set of competencies that change throughout life. To become proficient readers in a digital world, students need a solid foundation in reading, but also the ability to think critically, and adjust their behavior according to the task, as well as to motivate themselves to persevere in the face of difficulties.

In this sense, Spain could benefit from strengthening students' knowledge and skills to navigate through ambiguity, and to contrast and validate points of view. However, adding specific subjects on digital skills in school without adjusting other parts of the curriculum could be problematic. It is important to balance content and competencies to address new societal demands without overloading the curriculum. The challenge is to try to respond to the changing needs while minimizing the expansion and overload of content. One possibility for finding a balance between curricular updating and overload is to incorporate cross-cutting themes or competencies into existing subjects (OECD, 2020), as well as the use of innovative teaching and learning methods (Paniagua & Istance, 2018; Pérez et al., 2018).

TECHNOLOGY IN EDUCATIONAL ASSESSMENT: ADAPTIVE TESTING AND PROCESS DATA

Definitions of school competencies, not only for reading as described above, but also for other competencies such as mathematics and science, are continually being updated to reflect changes in technology, society, economy, and culture (Fraillon et al., 2019; Mullis & Martin, 2019; OECD, 2018). Navigation has become a cross-cutting component of any kind of learning in a digital world and, as we will describe below, also of the way these competencies are assessed.

Large-scale assessments coordinated by international organizations such as PISA, TIMSS (Trends in International Mathematics and Science Study), or PIRLS (Progress in International Reading Literacy Study) are mostly administered by computer, and this is the main method of administration in Spain as well. Spanish public administrations have been trialing computerized tests for some time in their diagnostic or system evaluations, although their use is not yet widespread. The reality in Spanish classrooms is not dissimilar. Although common in many schools, computerized assessments in the classroom (exams, rubrics, etc.) have not yet become the status quo. However, the trend is clear: as access to the Internet and digital devices increases so does their use. Advances in computer technology have also substantially influenced the ways in which tests are conducted, administered, scored, and reported to test takers (Zenisky & Sireci, 2002). In this article, we focus on two emerging areas whose unstoppable growth has generated a paradigm shift in educational assessment. For more information on other applications such as computational psychometrics and machine learning, see for example von Davier et al. (2019); in this monograph, Elosua (2022).

The first and perhaps the most widespread is the possibility of computerized adaptive testing (Olea et al., 2010; in this monograph, Abad et al., 2022). Computerized adaptive tests allow us to increase measurement accuracy by using fewer questions or items per student. This is achieved by presenting students with items that are aligned with

[2] Question 3 of the Rapa Nui unit assesses the ability to distinguish fact from opinion. It is a partial credit item where the null response is scored 0, the partial response is scored 0.5, and the complete response is scored 1. The estimated percentage correct for the complete response on this item is less than 41% in Spain and less than 47% on average in OECD countries. Question 3 of Rapa Nui is a level 5 item. This means that students must have a proficiency level of 5 to have a 62% chance of getting the full score on this item.

their proficiency level. Unlike non-adaptive assessments that typically focus on assessing students of average performance, adaptive tests also allow for a more refined differentiation of student ability at the high and low ends of the student's proficiency level (Table 1). This aspect of adaptive testing is particularly relevant when comparing high- and low-achieving groups or students from advantaged and disadvantaged families. At the same time, the testing experience for students is better, as they do not need to answer questions that are either too difficult for them or too easy, which could cause their interest in the test to be reduced. Adaptive testing is already used in the vast majority of large-scale educational assessments such as PISA, TIMSS, and PIRLS, and it is even used in personality assessment (Pedrosa et al., 2016; Postigo et al., 2020; in this monograph, Abad et al., 2022).

The second revolutionary change in computerized assessment is the possibility of storing log files, also known as process data. These data contain information about the actions performed by examinees when interacting with the tasks presented to them on the computer and the time spent on each action during the process (Table 2). This type of data provides additional information beyond the response data, which usually shows whether the question was answered correctly or not (He et al., 2019, 2021; von Davier et al., 2019). For example, the amount of time that students spend answering tests may reflect their level of engagement (whether they are doing their best) and, consequently, may affect their performance on the test (Wise et al., 2021). This information can help policymakers, researchers, and educators better understand students' cognitive strategies and the underlying causes of low and high performance. This, in turn, can result in improved assessment design and lead to more effective training and learning programs (OECD, 2019a).

TABLE 1
SUMMARY OF THE ADVANTAGES AND DISADVANTAGES OF COMPUTERIZED ADAPTIVE TESTING COMPARED TO USING CONVENTIONAL TESTS

Advantages	Disadvantages
More efficient, improves measurement without increasing assessment time.	Requires more sophisticated methods of analysis (item response theory).
More effective, it optimizes measurement accuracy across a wider range of examinee proficiency (e.g., high and low achievers).	Requires larger question banks (especially in computerized adaptive tests at item level, somewhat less so in multistage adaptive format designs).
More equitable, greater uniformity in the accuracy of the measure across different proficiency levels of examinees allows for better informed decisions (e.g., when comparing students from advantaged and disadvantaged backgrounds).	Requires larger participant samples (e.g., PISA uses a minimum of 200 responses per item).
Better experience, no questions are presented that are either too difficult or too easy.	Requires computer-based administration which may pose a barrier for some students (the mode of test administration must not favor certain groups over others).

ILLUSTRATIVE EXAMPLE OF THE USE OF PROCESS DATA IN PISA

PISA is a triennial survey of 15-year-old students around the world that assesses the extent to which they have acquired the key knowledge and skills essential for full participation in societies. PISA introduced computerized adaptive testing for the first time in 2018 to assess the reading proficiency of 15-year-old students (the latest data available to date). The interactive nature of computerized adaptive assessments such as PISA makes them ideal candidates for analyses based on process data (Goldhammer et al., 2016; Vörös et al., 2021).

Students' ability to think, monitor, and adjust their activity to a particular task are essential aspects of reading in digital environments (OECD, 2019b). Digital readers not only need to follow linear information structures, but also construct their own texts by selecting and evaluating information from different sources. Good navigation, therefore, should be consistent with these goals. The PISA 2018 reading test allows students to navigate through the different tasks in such a way that the student can decide how to interact with the text and which strategy is most effective. For example, one student may decide to strictly follow the instructions for each task as they appear while others may prefer to explore the tasks that will come next in order to manage their time and knowledge differently.

The recent PISA report, *21st Century Readers: Developing Literacy Skills in a Digital World*, uses process data to group students based on their navigation of the scenario-based reading unit, Rapa Nui[3] (CR551, see Appendix A):

TABLE 2
SUMMARY OF NAVIGATION INDICATORS BASED ON PROCESS DATA

Quantity indicators	Quality indicators	Temporary indicators
Number of pages visited	Navigational behaviors and strategy	Time spent on the first page (median)
	Use of hyperlinks	Time spent on the first page (proportion)
		Effective transition rate between pages (more than three seconds)
		Time spent on the instruction page

Note: Adapted from (OECD, 2021a).

[3] The scenario-based reading unit, Rapa Nui (CR551) consists of three texts: a web page from a teacher's blog, a book review, and a news article from a science magazine. In these multiple-text reading situations, readers must make decisions about which of the available text fragments is the most important, relevant, accurate, and truthful.

- ✔ No navigation: students who have no navigation activities in either single-source text items (see Appendix A items 1, 2, 3, 4, and 5) or multiple-source text items (see Appendix A items 6 and 7)
- ✔ Limited navigation: students who navigate simply through items with single-source text (see Appendix A items 1, 2, 3, 4, and 5), but not through items with multiple-source text (see Appendix A items 6 and 7)
- ✔ Strictly focused navigation: students who strictly followed the task instructions to actively navigate items with text from multiple sources of information (see Appendix A items 6 and 7), but had limited navigation on items with single-source text (see Appendix A items 1, 2, 3, 4, and 5)
- ✔ Active exploratory navigation: students who actively navigate items with single and multiple source texts (see Appendix A items 1, 2, 3, 4, 5, 6, and 7).

Figure 1 shows that reading achievement is strongly related to students' type of navigation. A consistent pattern between reading scores and navigation patterns was found across most countries and economies. The data show a difference of 66 points was found between students who actively navigated between pages and those who did not engage in navigational activities. The 11% of students who were in the active exploratory navigating group had the highest reading scores. These students actively navigated through single-source (see Appendix A items 1, 2, 3, 4, and 5) and multiple-source (see Appendix A items 6, and 7) tasks. Their navigation in single-source text tasks exceeded the number of pages required to complete the reading task. That is, they read not only the required page, but also other accessible pages. This type of navigation could help students obtain an overview of the entire reading test and better prepare for the subsequent tasks. This could also explain, at least in part, why the active navigating group of students scored better on the reading test than the group of students who strictly followed the task instructions.

Girls presented better reading strategies than boys in all three indices included in PISA 2018 (understanding and remembering a text, writing a summary, and evaluating the credibility of information sources). This is repeated across all navigation groups. Even so, both boys and girls who actively navigated and explored the Rapa Nui unit had better knowledge of reading strategies for assessing the credibility of sources than students with limited or no navigation. In addition, gender differences in reading strategies are narrower among students with more active navigating than those with no navigating (Figure 2).

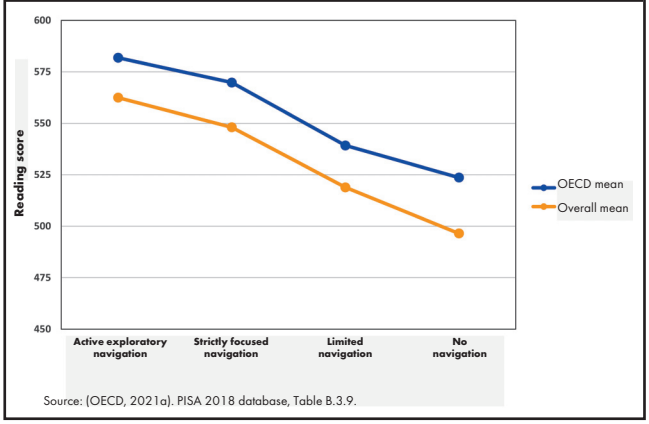
DISCUSSION

The use of technology in educational assessment enables the use of new data sources for both the assessment process (test design) and the product (test scores). Computerized adaptive testing and process data are two of the emerging areas of greatest benefit in educational assessment. The use of process data in computerized testing is particularly valuable when assessing competencies that require interactive tasks and the use of technology. In the case of reading proficiency assessment, the example discussed in this article, students need to use information and communication technologies (ICT) to access texts through search engines, use links and tabs, process information from multiple sources, evaluate the quality of information sources, detect possible conflicts, and resolve them. Process data, more specifically the sequences of actions performed by students when responding to the test, allow us to identify the navigation strategies used by students when interacting with texts in a digital environment. These data could be used to measure the degree to which students engage with the task in the expected way and thus improve the validity of the interpretations, but they could also be used alongside the students' responses to calculate their scores on the competencies tested (Wise et al., 2021).

In this way, process data, which are a priori inherent aspects of the assessment method and test design (the process), become to some extent the object of assessment (the product). In other words, technology has permeated all layers of educational assessment in such a way that it is almost impossible to disentangle the measurement from the construct. The possibilities that this type of data has for educational evaluation are countless (Jiao et al., 2021), but as the proverb reminds us: it is important not to put the cart before the horse. There are at least three important considerations for optimizing the use of this type of data:

Design the test with the objective of extracting and using process data. Process data are a by-product of inherent features of the software used to perform the assessment. That is, in the vast majority of cases, tests are not constructed and designed with the goal of using these data, but rather they are used *serendipitously* to improve the process and product of the assessment once the test has been administered and the answers to the questions coded. While there is nothing wrong with improving the assessment process and product—indeed, the standards for educational and psychological testing recommend that this should be done (AERA et al., 2014)—it is essential to consider the implications of such improvements for other important areas of the assessment. For example, if the goal is to improve the product of the assessments (the scores) but the use of process data has not been taken into account in the test design, the content validity (Pedrosa et al., 2014) of the test may be altered with respect to its conceptual assessment framework. In other words, the risk of giving meaning to data *a posteriori* is that it may divert the purpose and content of the assessment from that originally proposed

FIGURE 1
ASSOCIATION BETWEEN READING PERFORMANCE AND NAVIGATING BEHAVIORS



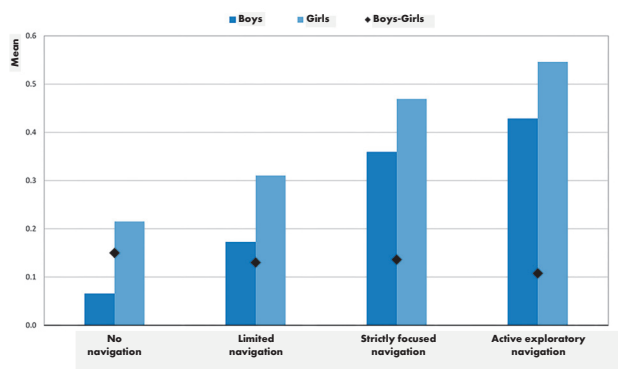
Source: (OECD, 2021a). PISA 2018 database, Table B.3.9.

in the conceptual assessment frameworks. For a discussion of this issue see, for example, how the PISA rankings would change if this type of data were taken into account in the calculation of scores (Pohl et al., 2021).

The extraction of process data should be based on theoretical and analytical considerations. The selection of the information that is recorded while the examinee is answering the test does not respond to theoretical or analytical considerations. In other words, the software blindly records all the examinee's activity, whether it is relevant or not. Therefore, the extraction of process data must be guided by theoretical and analytical considerations to avoid spurious interpretations. A particularly interesting application of using process data to improve the measurement (the process) is to use response times to identify differences in the response process between different cultural and linguistic groups and thus improve the validity of interpretations that might otherwise be overlooked in differential item functioning[4] (Ercikan et al., 2020).

The interpretation of process data should be guided by cognitive models. It is important to note that cognitive response processes are not observable in either conventional or digital tests. As with other methods for obtaining validity evidence for response processes, such as cognitive labs or think-aloud protocols (Padilla & Benítez, 2014), the log files do not reflect the cognitive processes per se, but rather traces of the cognitive processes that the students used. Their interpretation, therefore, requires the use of mixed methods that combine qualitative and quantitative interpretations. In other words, it is essential that the interpretation of process data be guided by theoretical models, in this case, cognitive ones. For example, the assessment of complex constructs such as reasoning or critical thinking may involve cognitive processes such as response times explicitly included in the definition of the construct.

FIGURE 2
ASSOCIATION BETWEEN READING STRATEGIES FOR ASSESSING THE CREDIBILITY OF INFORMATION SOURCES AND NAVIGATING BEHAVIORS



Note: All differences between girls and boys are statistically significant.
Source: (OECD, 2021a). PISA 2018 database, Table B.5.29.

Another example, as we saw in this article, is the inclusion of navigation behaviors as a cross-cutting aspect of the learning of reading literacy in a digital world.

Technology enhances knowledge to unimaginable levels, whether this knowledge is adequate or not. In fact, it is common for inadequate or erroneous knowledge to have a greater capacity for expansion. See as an example the transmission of information via social networks; a false news item spreads at a faster rate than a true one (Vosoughi et al., 2018). *Mutatis mutandis,* studies that are the most difficult to replicate tend to be the most cited (Serra-Garcia & Gneezy, 2021). In Spain, moreover, the use of tests in the professional environment is enormous, and training and knowledge about their use is relatively low (Muñiz et al., 2020; in this monograph, Hernández et al., 2022). It is the researcher's responsibility to ensure that the use of data and technology is appropriate for the purposes of the assessment and serves the people involved in a reliable, valid, and fair way, but it is also the user's responsibility to know when, how, and for what purpose to use the data. We hope that this article will encourage reflection on both aspects.

CONFLICT OF INTEREST

There is no conflict of interest.

REFERENCES

Abad, F. J., Schames, R., Sorrel, M., Nájera, P., García-Garzón, E., Garrido, L. E., & Jiménez, M. (2022). Construyendo tests adaptativos de elección forzosa "on the fly" para la medición de la personalidad [Building Adaptive Forced-Choice Tests "On The Fly" For Personality Measurement]. *Papeles del Psicólogo, 43*(1), 29-35. https://doi.org/10.23923/pap.psicol.2982

AERA, APA, & NCME. (2014). *Standards for educational and psychological testing (2nd ed.)*. Washington, DC: American Educational Research Association (AERA).

Andrés, J. C., Aguado, D., & de Miguel, J. (2022). ¿Qué hay detras de LinkedIn? Midiendo a través de rúbricas las LinkedIn Big Four Dimensions [What's behind LinkedIn? Measuring the LinkedIn Big Four Dimensions through Rubrics]. *Papeles del Psicólogo, 43*(1), 12-20. https://doi.org/10.23923/pap.psicol.2979

Breakstone, J., McGrew, S., Smith, M., Ortega, T., & Wineburg, S. (2018). Why we need a new approach to teaching digital literacy. *Phi Delta Kappan, 99*(6), 27–32. https://doi.org/10.1177/0031721718762419

Clinton, V. (2019). Reading from paper compared to screens: A systematic review and meta-analysis. *Journal of Research in Reading, 42*(2), 288–325. https://doi.org/10.1111/1467-9817.12269

Elosua, P. (2022). Impacto de la TIC en el entorno evaluativo. Innovaciones al servicio de la mejora continua [Impact of ICT on the assessment environment. Innovations for the purpose of continuous improvement]. *Papeles del Psicólogo, 43*(1), 3-11. https://doi.org/10.23923/pap.psicol.2985

[4] The existence of differential item functioning (DIF) indicates that people who are equally skilled in the assessed competency have different probabilities of responding correctly to an item, and this is related to bias, i.e., favoring one group over another in the assessment.

Delgado, P., Vargas, C., Ackerman, R., & Salmerón, L. (2018). Don't throw away your printed books: A meta-analysis on the effects of reading media on reading comprehension. *Educational Research Review, 25*, 23–38. https://doi.org/10.1016/j.edurev.2018.09.003

Fonseca-Pedrero, E., Ródenas, Gabriel, Pérez-Albéniz, A., Al-Halabí, S., Pérez, M., & Muñiz, J. (2022). La hora de la evaluación ambulatoria [The Time of Ambulatory Assessment]. *Papeles del Psicólogo, 43*(1), 21-28 https://doi.org/10.23923/pap.psicol.2983

Ercikan, K., Guo, H., & He, Q. (2020). Use of response process data to inform group comparisons and fairness research. *Educational Assessment, 25*(3), 179–197. https://doi.org/10.1080/10627197.2020.1804353

Fraillon, J., Ainley, J., Schulz, W., Duckworth, D., & Friedman, T. (2019). IEA International Computer and Information Literacy Study 2018 Assessment Framework. *IEA International Computer and Information Literacy Study 2018 Assessment Framework*. https://doi.org/10.1007/978-3-030-19389-8

Goldhammer, F., Martens, T., Christoph, G., & Lüdtke, O. (2016). *Test-taking engagement in PIAAC. OECD Education Working Papers, No. 133, OECD Publishing, Paris, https://doi.org/10.1787/5jlzfl6fhxs2-en*

He, Q., Borgonovi, F., & Paccagnella, M. (2019). *Using process data to understand adults' problem-solving behaviour in the Programme for the International Assessment of Adult Competencies (PIAAC): Identifying generalised patterns across multiple tasks with sequence mining*. OECD Publishing, Paris. https://www.oecd-ilibrary.org/education/using-process-data-to-understand-adults-problem-solving-behaviour-in-the-programme-for-the-international-assessment-of-adult-competencies-piaac_650918f2-en

He, Q., Borgonovi, F., & Paccagnella, M. (2021). Leveraging process data to assess adults' problem-solving skills: Using sequence mining to identify behavioral patterns across digital tasks. *Computers & Education, 166*, 104170.

Hernández, A., Elosua, P., Fernández-Hermida, J. R., & Muñiz, J. (2022). Comisión de Test: Veinticinco años velando por la calidad de los test [Test Commission: Twenty-five years working on test quality]. Papeles del Psicólogo, 43(1), 55-62. https://doi.org/10.23923/pap.psicol.2978

INEE. (2019). PISA 2018. Items liberados. Secretaria General Técnica. Subdirección General de Atención al Ciudadano, Documentación y Publicaciones [PISA 2018. Items released. General Technical Secretary. Subdirectorate General for Citizen Services, Documentation and Publications]. https://sede.educacion.gob.es/ publiventa/descarga.action?f_codigo_agc=20232

INEE. (2021). Nota país: Lectores del siglo XXI. PISA 2018. Secretaria General Tecnica. Subdireccion General de Atencion al Ciudadano, Documentacion y Publicaciones [Country note: Readers of the 21st century. PISA 2018. Technical General Secretary. Subdirectorate General for Citizen Services, Documentation, and Publications]. https://doi.org/10.1016/j.compedu.2021.104170

Jiao, H., He, Q., & Veldkamp, B. P. (2021). Editorial: Process data in educational and psychological measurement. *Frontiers in Psychology, 12*. https://doi.org/10.3389/fpsyg.2021.793399

Macedo-Rouet, M., Salmerón, L., Ros, C., Pérez, A., Stadtler, M., & Rouet, J.-F. (2020). Are frequent users of social network sites good information evaluators? An investigation of adolescents' sourcing abilities. *Journal for the Study of Education and Development, 43*(1), 101–138. https://doi.org/10.1080/02103702.2019.1690849

McGrew, S., Breakstone, J., Ortega, T., Smith, M., & Wineburg, S. (2018). Can students evaluate online sources? Learning from assessments of civic online reasoning. *Theory & Research in Social Education, 46*(2), 165–193. https://doi.org/10.1080/00933104.2017.1416320

Mullis, I. V. S., & Martin, M. O. (2019). *Assessment Frameworks PIRLS 2021*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Muñiz, J., Hernández, A., & Fernández-Hermida, J. R. (2020). Utilización de los test en España: el punto de vista de los psicólogos [Test use in Spain: the psychologists' viewpoint]. *Papeles del Psicólogo, 41*(1), 1-15. https://doi.org/10.23923/pap.psicol2020.2921

OECD. (2011). *PISA 2009 Results: Students On Line*. PISA, OECD Publishing, Paris. https://doi.org/10.1787/9789264112995-en

OECD. (2018). *PISA 2021 Mathematics Framework*. PISA, OECD Publishing, Paris.

OECD. (2019a). *Beyond proficiency: Using log files to understand respondent behaviour in the survey of adult skills*. OECD Publishing, Paris. https://doi.org/10.1787/0b1414ed-en

OECD. (2019b). *PISA 2018 Assessment and analytical framework*. PISA, OECD Publishing, Paris. https://doi.org/10.1787/b25efab8-en

OECD. (2020). *Curriculum Overload*. PISA, OECD Publishing, Paris. https://doi.org/10.1787/3081ceca-en

OECD. (2021a). *21st-Century readers: Developing literacy skills in a digital world*. PISA, OECD Publishing, Paris. https://doi.org/10.1787/a83d84cb-en

INEE. (2021b). *Nota país: Lectores del siglo XXI. PISA 2018*. Secretaria General Técnica. Subdirección General de Atención al Ciudadano, Documentación y Publicaciones [Country note: Readers of the 21st century. PISA 2018. General Technical Secretary. Subdirectorate General for Citizen Services, Documentation and Publications].

Olea, J., Abad, F. J., & Barrada, J. R. (2010). Tests informatizados y otros nuevos tipos de tests [Computerized tests and other new types of test]. *Papeles del Psicólogo, 31*(1), 94–107.

Padilla, J. L., & Benítez, I. (2014). Evidencia de validez basada en los procesos de respuesta [Validity evidence based on response processes]. *Psicothema, 26*(1), 136–144. https://doi.org/10.7334/PSICOTHEMA2013.259

Paniagua, A., & Istance, D. (2018). *Teachers as designers of learning environments*. OECD Publishing, Paris. https://doi.org/10.1787/9789264085374-en

Pedrosa, I., Suárez-Álvarez, J., & García-Cueto, E. (2014). Evidencias sobre la validez de contenido: Avances teóricos y métodos para su estimación [Evidence of content validity: Theoretical advances and estimation methods]. *Acción Psicológica, 10*(2), 3. https://doi.org/10.5944/ap.10.2.11820

Pedrosa, I., Suárez-Álvarez, J., García-Cueto, E., & Muñiz, J. (2016). A computerized adaptive test for enterprising personality assessment in youth. *Psicothema, 28*(4). https://doi.org/10.7334/psicothema2016.68

Pérez, A., Potocki, A., Stadtler, M., Macedo-Rouet, M., Paul, J., Salmerón, L.,

& Rouet, J.-F. (2018). Fostering teenagers' assessment of information reliability: Effects of a classroom intervention focused on critical source dimensions. *Learning and Instruction, 58*, 53–64. https://doi.org/10.1016/j.learninstruc.2018.04.006

Pohl, S., Ulitzsch, E., & von Davier, M. (2021). Reframing rankings in educational assessments. *Science, 372*(6540), 338–340. https://doi.org/10.1126/science.abd3300

Postigo, Á., Cuesta, M., Pedrosa, I., Muñiz, J., & García-Cueto, E. (2020). Development of a computerized adaptive test to assess entrepreneurial personality. *Psicologia: Reflexão e Crítica, 33*(1), 6. https://doi.org/10.1186/s41155-020-00144-x

Santamaría, P. & Sánchez-Sánchez, F. (2022). Cuestiones abiertas en el uso de las nuevas tecnologías en la evaluación psicológica [Open questions in the use of new technologies in psychological assessment]. *Papeles del Psicólogo, 43*(1), 48-54. https://doi.org/10.23923/pap.psicol.2984

Serra-Garcia, M., & Gneezy, U. (2021). Nonreplicable publications are cited more than replicable ones. *Science Advances, 7*(21). https://doi.org/10.1126/SCIADV.ABD1705/SUPPL_FILE/SCIADV.ABD1705_SM.PDF

Suarez-Alvarez, J. (2021). *Are 15-year-olds prepared to deal with fake news and misinformation?* PISA, OECD Publishing, Paris.

von Davier, A. A., Deonovic, B., Yudelson, M., Polyak, S. T., & Woo, A. (2019). Computational psychometrics approach to holistic learning and assessment systems. *Frontiers in Education, 4*. https://doi.org/10.3389/feduc.2019.00069

von Davier, M., Khorramdel, L., He, Q., Shin, H. J., & Chen, H. (2019). Developments in psychometric population models for technology-based large-scale assessments: An overview of challenges and opportunities. *Journal of Educational and Behavioral Statistics, 44*(6), 671–705. https://doi.org/10.3102/1076998619881789

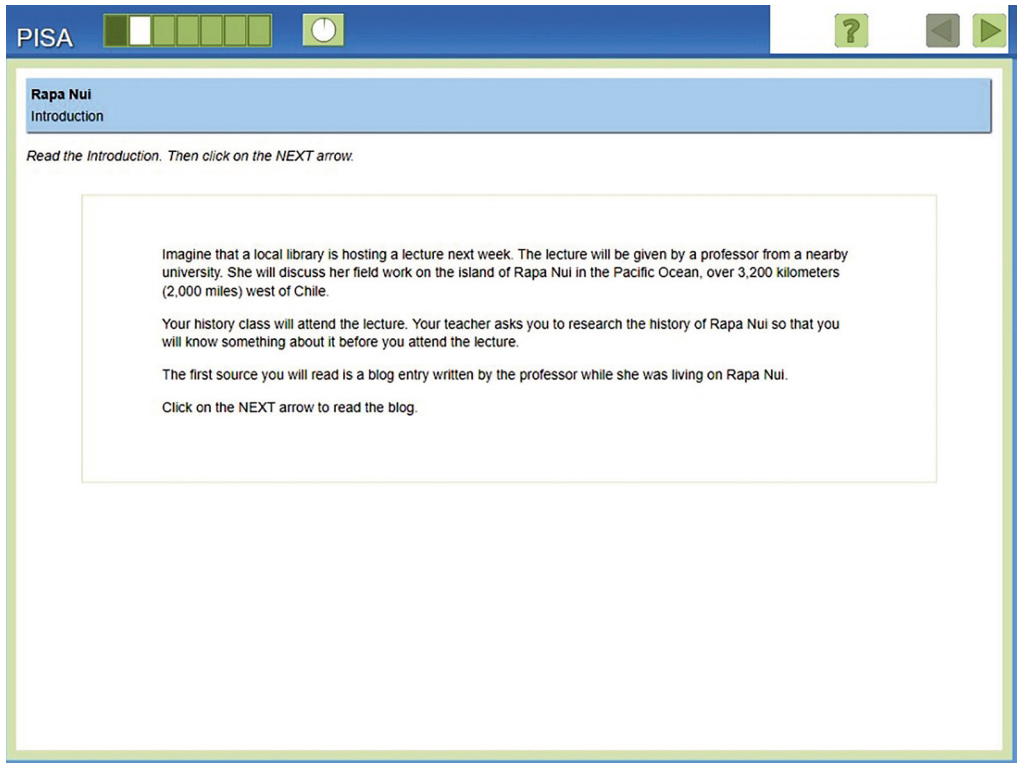
Vörös, Z., Kehl, D., & Rouet, J.-F. (2021). Task characteristics as source of difficulty and moderators of the effect of time-on-task in digital problem-solving. *Journal of Educational Computing Research, 58*(8), 1494–1514. https://doi.org/10.1177/0735633120945930

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science, 359*(6380), 1146–1151. https://doi.org/10.1126/SCIENCE.AAP9559/SUPPL_FILE/AAP9559_VOSOUGHI_SM.PDF

Wise, S. L., Im, S., & Lee, J. (2021). The Impact of Disengaged Test Taking on a State's Accountability Test Results. *Educational Assessment, 26*(3), 163–174. https://doi.org/10.1080/10627197.2021.1956897

Zenisky, A., & Sireci, S. (2002). Technological Innovations in Large-Scale Assessment. *Applied Measurement in Education, 15*(4), 337–362.

Appendix A
Unit CR551: Easter Island (INEE, 2019)



PISA

Rapa Nui
Question 1 / 7

Refer to the Professor's Blog on the right. Click on a choice to answer the question.

According to the blog, when did the professor start her field work?

- ○ During the 1990s
- ○ Nine months ago
- ○ One year ago
- ○ At the beginning of May

Blog

www.theprofessorblog.com/fieldwork/RapaNui

The Professor's Blog

Posted May 23, 11:22 a.m.

As I look out of my window this morning, I see the landscape I have learned to love here on Rapa Nui, which is known in some places by the name Easter Island. The grasses and shrubs are green, the sky is blue, and the old, now extinct volcanoes rise up in the background.

I am a bit sad knowing that this is my last week on the island. I have finished my field work and will be returning home. Later today, I will take a walk through the hills and say goodbye to the *moai* that I have been studying for the past nine months. Here is a picture of some of these massive statues.

If you have been following my blog this year, then you know that the people of Rapa Nui carved these *moai* hundreds of years ago. These impressive *moai* had been carved in a single quarry on the eastern part of the island. Some of them weighed thousands of pounds, yet the people of Rapa Nui were able to move them to locations far away from the quarry without cranes or any heavy equipment.

For years, archeologists did not know how these massive statues were moved. It remained a mystery until the 1990s, when a team of archeologists and residents of Rapa Nui demonstrated that the *moai* could have been transported and raised using ropes made from plants and wooden rollers and tracks made from large trees that had once thrived on the island. The mystery of the *moai* was solved.

Another mystery remained, however. What happened to these plants and large trees that had been used to move the *moai*? As I said, when I look out of my window, I see grasses and shrubs and a small tree or two, but nothing that could have been used to move these huge statues. It is a fascinating puzzle, one that I will explore in future posts and lectures. Until then, you may wish to investigate the mystery yourself. I suggest you begin with a book called *Collapse* by Jared Diamond. This review of *Collapse* is a good place to start.

Traveler_14 — *May 24, 4:31 p.m.*
Hi Professor! I love following your work on Easter Island. I can't wait to check out *Collapse*!

KB_Island — *May 25, 9:07 a.m.*
I also love reading about your experiences on Easter Island, however, I think there is another theory that should be considered. Check out this article:
www.sciencenews.com/Polynesian_rats_Rapa_Nui

PISA

Rapa Nui
Question 2 / 7

Refer to the Professor's Blog on the right. Type your answer to the question.

In the last paragraph of the blog, the professor writes: "Another mystery remained..."

To what mystery does she refer?

Blog

www.theprofessorblog.com/fieldwork/RapaNui

The Professor's Blog

Posted May 23, 11:22 a.m.

As I look out of my window this morning, I see the landscape I have learned to love here on Rapa Nui, which is known in some places by the name Easter Island. The grasses and shrubs are green, the sky is blue, and the old, now extinct volcanoes rise up in the background.

I am a bit sad knowing that this is my last week on the island. I have finished my field work and will be returning home. Later today, I will take a walk through the hills and say goodbye to the *moai* that I have been studying for the past nine months. Here is a picture of some of these massive statues.

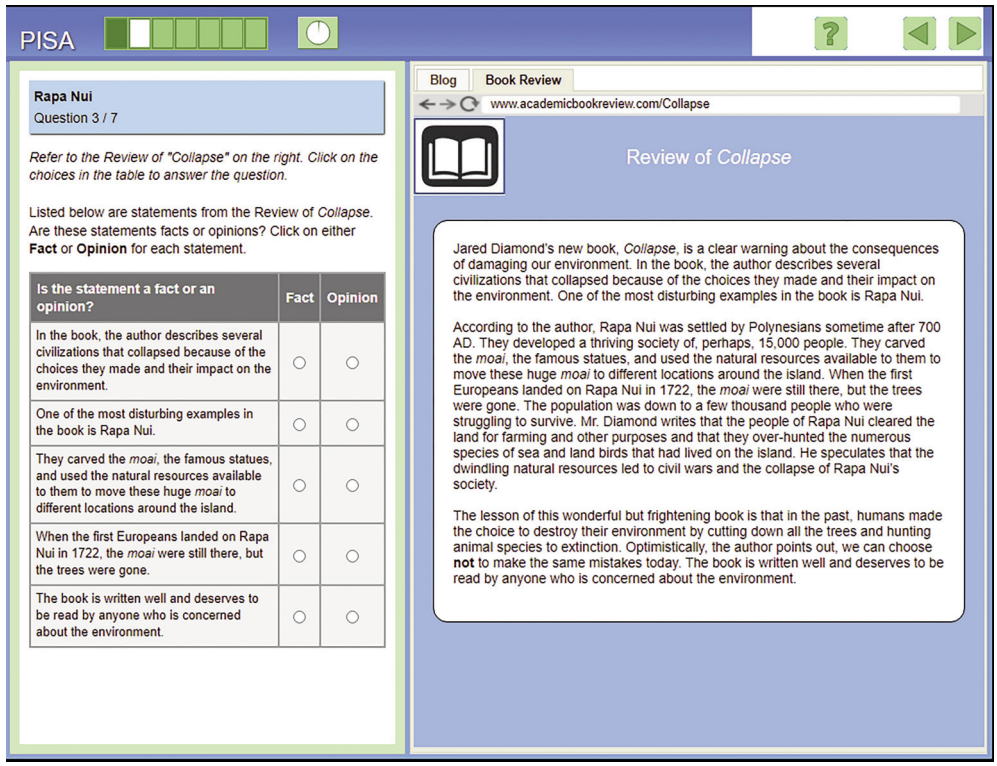
If you have been following my blog this year, then you know that the people of Rapa Nui carved these *moai* hundreds of years ago. These impressive *moai* had been carved in a single quarry on the eastern part of the island. Some of them weighed thousands of pounds, yet the people of Rapa Nui were able to move them to locations far away from the quarry without cranes or any heavy equipment.

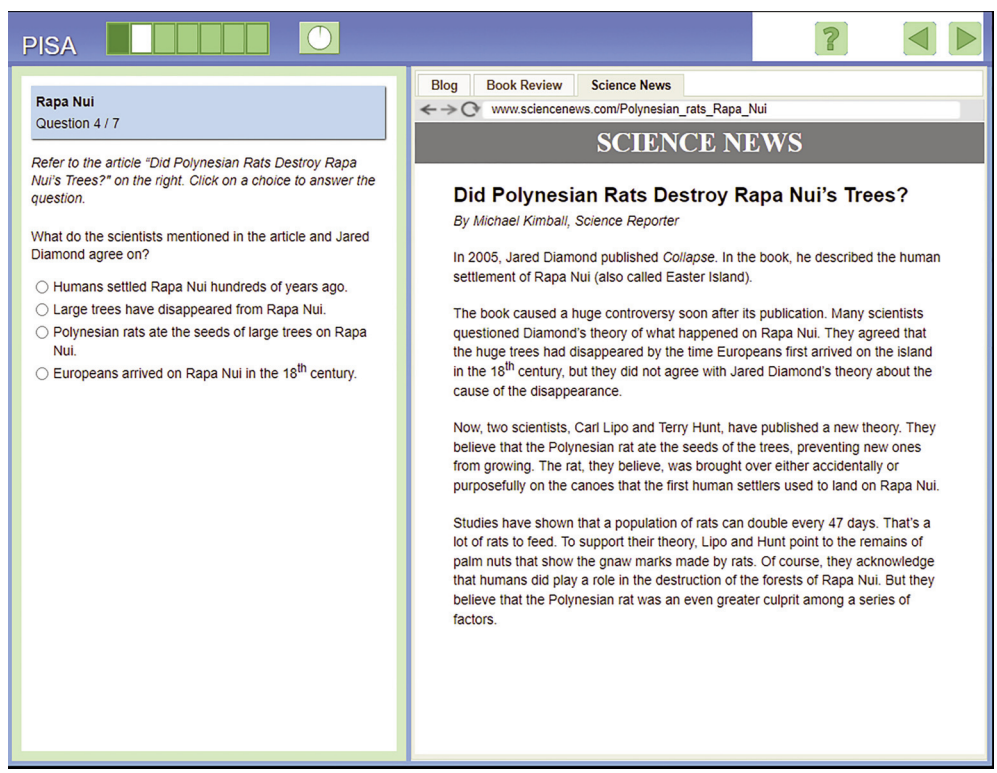
For years, archeologists did not know how these massive statues were moved. It remained a mystery until the 1990s, when a team of archeologists and residents of Rapa Nui demonstrated that the *moai* could have been transported and raised using ropes made from plants and wooden rollers and tracks made from large trees that had once thrived on the island. The mystery of the *moai* was solved.

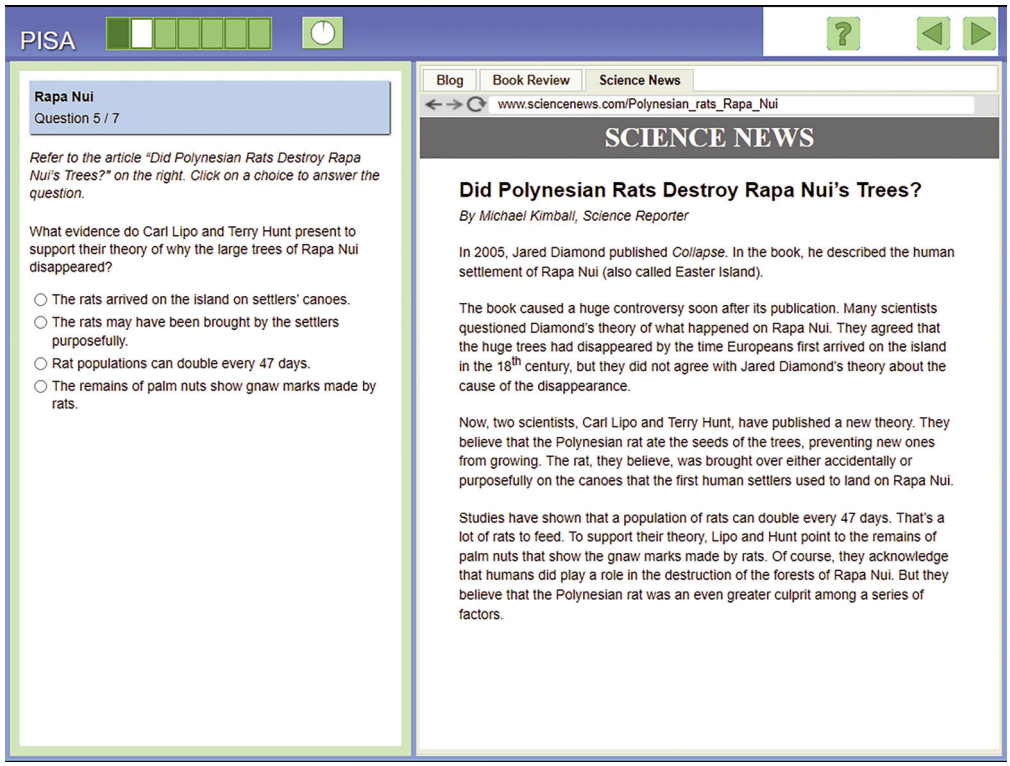
Another mystery remained, however. What happened to these plants and large trees that had been used to move the *moai*? As I said, when I look out of my window, I see grasses and shrubs and a small tree or two, but nothing that could have been used to move these huge statues. It is a fascinating puzzle, one that I will explore in future posts and lectures. Until then, you may wish to investigate the mystery yourself. I suggest you begin with a book called *Collapse* by Jared Diamond. This review of *Collapse* is a good place to start.

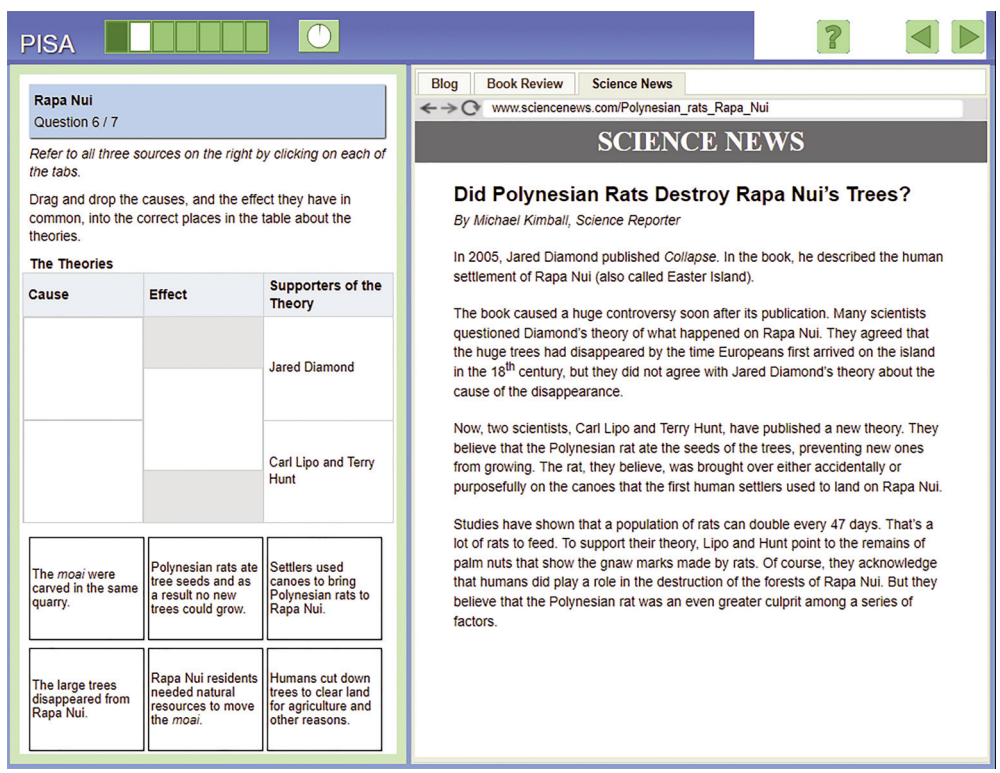
Traveler_14 — *May 24, 4:31 p.m.*
Hi Professor! I love following your work on Easter Island. I can't wait to check out *Collapse*!

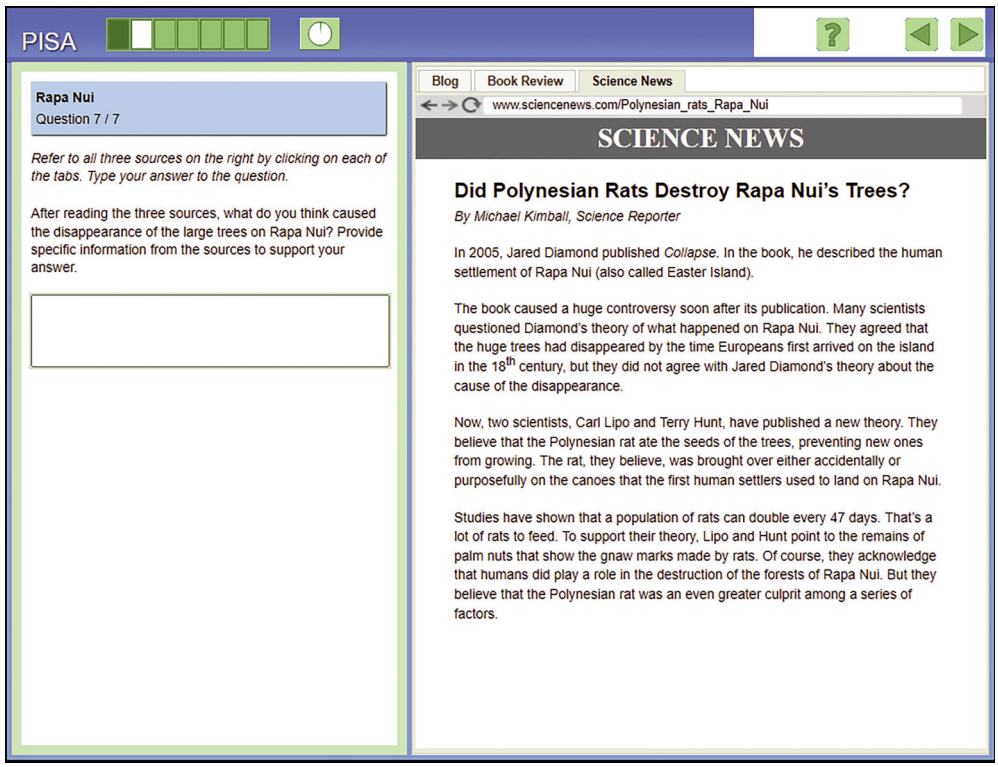
KB_Island — *May 25, 9:07 a.m.*
I also love reading about your experiences on Easter Island, however, I think there is another theory that should be considered. Check out this article:
www.sciencenews.com/Polynesian_rats_Rapa_Nui


PISA
Rapa Nui
Question 3 / 7
Refer to the Review of "Collapse" on the right. Click on the choices in the table to answer the question.
Listed below are statements from the Review of Collapse. Are these statements facts or opinions? Click on either Fact or Opinion for each statement.
Is the statement a fact or an opinion?
Fact
Opinion
In the book, the author describes several civilizations that collapsed because of the choices they made and their impact on the environment.
One of the most disturbing examples in the book is Rapa Nui.
They carved the moai, the famous statues, and used the natural resources available to them to move these huge moai to different locations around the island.
When the first Europeans landed on Rapa Nui in 1722, the moai were still there, but the trees were gone.
The book is written well and deserves to be read by anyone who is concerned about the environment.
Blog
Book Review
www.academicbookreview.com/Collapse
Review of Collapse
Jared Diamond's new book, Collapse, is a clear warning about the consequences of damaging our environment. In the book, the author describes several civilizations that collapsed because of the choices they made and their impact on the environment. One of the most disturbing examples in the book is Rapa Nui.
According to the author, Rapa Nui was settled by Polynesians sometime after 700 AD. They developed a thriving society of, perhaps, 15,000 people. They carved the moai, the famous statues, and used the natural resources available to them to move these huge moai to different locations around the island. When the first Europeans landed on Rapa Nui in 1722, the moai were still there, but the trees were gone. The population was down to a few thousand people who were struggling to survive. Mr. Diamond writes that the people of Rapa Nui cleared the land for farming and other purposes and that they over-hunted the numerous species of sea and land birds that had lived on the island. He speculates that the dwindling natural resources led to civil wars and the collapse of Rapa Nui's society.
The lesson of this wonderful but frightening book is that in the past, humans made the choice to destroy their environment by cutting down all the trees and hunting animal species to extinction. Optimistically, the author points out, we can choose not to make the same mistakes today. The book is written well and deserves to be read by anyone who is concerned about the environment.


PISA
Rapa Nui
Question 4 / 7
Refer to the article "Did Polynesian Rats Destroy Rapa Nui's Trees?" on the right. Click on a choice to answer the question.
What do the scientists mentioned in the article and Jared Diamond agree on?
Humans settled Rapa Nui hundreds of years ago.
Large trees have disappeared from Rapa Nui.
Polynesian rats ate the seeds of large trees on Rapa Nui.
Europeans arrived on Rapa Nui in the 18th century.
Blog
Book Review
Science News
www.sciencenews.com/Polynesian_rats_Rapa_Nui
SCIENCE NEWS
Did Polynesian Rats Destroy Rapa Nui's Trees?
By Michael Kimball, Science Reporter
In 2005, Jared Diamond published Collapse. In the book, he described the human settlement of Rapa Nui (also called Easter Island).
The book caused a huge controversy soon after its publication. Many scientists questioned Diamond's theory of what happened on Rapa Nui. They agreed that the huge trees had disappeared by the time Europeans first arrived on the island in the 18th century, but they did not agree with Jared Diamond's theory about the cause of the disappearance.
Now, two scientists, Carl Lipo and Terry Hunt, have published a new theory. They believe that the Polynesian rat ate the seeds of the trees, preventing new ones from growing. The rat, they believe, was brought over either accidentally or purposefully on the canoes that the first human settlers used to land on Rapa Nui.
Studies have shown that a population of rats can double every 47 days. That's a lot of rats to feed. To support their theory, Lipo and Hunt point to the remains of palm nuts that show the gnaw marks made by rats. Of course, they acknowledge that humans did play a role in the destruction of the forests of Rapa Nui. But they believe that the Polynesian rat was an even greater culprit among a series of factors.


PISA
Rapa Nui
Question 5 / 7
Refer to the article "Did Polynesian Rats Destroy Rapa Nui's Trees?" on the right. Click on a choice to answer the question.
What evidence do Carl Lipo and Terry Hunt present to support their theory of why the large trees of Rapa Nui disappeared?
The rats arrived on the island on settlers' canoes.
The rats may have been brought by the settlers purposefully.
Rat populations can double every 47 days.
The remains of palm nuts show gnaw marks made by rats.
Blog
Book Review
Science News
www.sciencenews.com/Polynesian_rats_Rapa_Nui
SCIENCE NEWS
Did Polynesian Rats Destroy Rapa Nui's Trees?
By Michael Kimball, Science Reporter
In 2005, Jared Diamond published Collapse. In the book, he described the human settlement of Rapa Nui (also called Easter Island).
The book caused a huge controversy soon after its publication. Many scientists questioned Diamond's theory of what happened on Rapa Nui. They agreed that the huge trees had disappeared by the time Europeans first arrived on the island in the 18th century, but they did not agree with Jared Diamond's theory about the cause of the disappearance.
Now, two scientists, Carl Lipo and Terry Hunt, have published a new theory. They believe that the Polynesian rat ate the seeds of the trees, preventing new ones from growing. The rat, they believe, was brought over either accidentally or purposefully on the canoes that the first human settlers used to land on Rapa Nui.
Studies have shown that a population of rats can double every 47 days. That's a lot of rats to feed. To support their theory, Lipo and Hunt point to the remains of palm nuts that show the gnaw marks made by rats. Of course, they acknowledge that humans did play a role in the destruction of the forests of Rapa Nui. But they believe that the Polynesian rat was an even greater culprit among a series of factors.


PISA
Rapa Nui
Question 6 / 7
Refer to all three sources on the right by clicking on each of the tabs.
Drag and drop the causes, and the effect they have in common, into the correct places in the table about the theories.
The Theories
Cause
Effect
Supporters of the Theory
Jared Diamond
Carl Lipo and Terry Hunt
The moai were carved in the same quarry.
Polynesian rats ate tree seeds and as a result no new trees could grow.
Settlers used canoes to bring Polynesian rats to Rapa Nui.
The large trees disappeared from Rapa Nui.
Rapa Nui residents needed natural resources to move the moai.
Humans cut down trees to clear land for agriculture and other reasons.
Blog
Book Review
Science News
www.sciencenews.com/Polynesian_rats_Rapa_Nui
SCIENCE NEWS
Did Polynesian Rats Destroy Rapa Nui's Trees?
By Michael Kimball, Science Reporter
In 2005, Jared Diamond published Collapse. In the book, he described the human settlement of Rapa Nui (also called Easter Island).
The book caused a huge controversy soon after its publication. Many scientists questioned Diamond's theory of what happened on Rapa Nui. They agreed that the huge trees had disappeared by the time Europeans first arrived on the island in the 18th century, but they did not agree with Jared Diamond's theory about the cause of the disappearance.
Now, two scientists, Carl Lipo and Terry Hunt, have published a new theory. They believe that the Polynesian rat ate the seeds of the trees, preventing new ones from growing. The rat, they believe, was brought over either accidentally or purposefully on the canoes that the first human settlers used to land on Rapa Nui.
Studies have shown that a population of rats can double every 47 days. That's a lot of rats to feed. To support their theory, Lipo and Hunt point to the remains of palm nuts that show the gnaw marks made by rats. Of course, they acknowledge that humans did play a role in the destruction of the forests of Rapa Nui. But they believe that the Polynesian rat was an even greater culprit among a series of factors.


PISA
Rapa Nui
Question 7 / 7
Refer to all three sources on the right by clicking on each of the tabs. Type your answer to the question.
After reading the three sources, what do you think caused the disappearance of the large trees on Rapa Nui? Provide specific information from the sources to support your answer.
Blog
Book Review
Science News
www.sciencenews.com/Polynesian_rats_Rapa_Nui
SCIENCE NEWS
Did Polynesian Rats Destroy Rapa Nui's Trees?
By Michael Kimball, Science Reporter
In 2005, Jared Diamond published Collapse. In the book, he described the human settlement of Rapa Nui (also called Easter Island).
The book caused a huge controversy soon after its publication. Many scientists questioned Diamond's theory of what happened on Rapa Nui. They agreed that the huge trees had disappeared by the time Europeans first arrived on the island in the 18th century, but they did not agree with Jared Diamond's theory about the cause of the disappearance.
Now, two scientists, Carl Lipo and Terry Hunt, have published a new theory. They believe that the Polynesian rat ate the seeds of the trees, preventing new ones from growing. The rat, they believe, was brought over either accidentally or purposefully on the canoes that the first human settlers used to land on Rapa Nui.
Studies have shown that a population of rats can double every 47 days. That's a lot of rats to feed. To support their theory, Lipo and Hunt point to the remains of palm nuts that show the gnaw marks made by rats. Of course, they acknowledge that humans did play a role in the destruction of the forests of Rapa Nui. But they believe that the Polynesian rat was an even greater culprit among a series of factors.