

Ring, Malte; Oberrauch, Luis

Working Paper

Measuring Economic Graph Competence

Suggested Citation: Ring, Malte; Oberrauch, Luis (2024) : Measuring Economic Graph Competence, ZBW - Leibniz Information Centre for Economics, Kiel, Hamburg

This Version is available at:

<https://hdl.handle.net/10419/300727>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Measuring Economic Graph Competence

Malte Ring and Luis Oberrauch

Malte Ring* (corresponding author), Tübingen University, D-72074 Tübingen, Germany, malte.ring@uni-tuebingen.de, ORCID 0000-0002-5248-0427

Luis Oberrauch, University of Kaiserslautern-Landau, D-76829 Landau, Germany, luis.oberrauch@rptu.de

Abstract

The ability to work with graphs is at the core of the economic domain and is also one of the central challenges for novices in the field. To accurately assess the graph competence of higher education students, we developed and tested an 18-item graph competence instrument with different economics graphs. The sample consisted of 579 students from multiple higher education institutions in southwestern Germany. Item analysis based on item response theory revealed that the instrument addresses a wide range of ability levels and discriminates sufficiently between low- and high-ability learners. Concerning content validity, we find meaningful correlations between test scores and domain-specific knowledge and generic understanding of numerical quantities. Item difficulties indicate that questions that go beyond simple graph operations are challenging for learners.

Keywords: graph competence, graph comprehension, economic competence, item response theory, measurement instrument, supply and demand

Working Paper, August 2024

Acknowledgement: We thank the university of Mainz and namely Prof. Dr. Zlatkin-Troitschanskaia and Prof. Dr. Roland Happ for the opportunity to use items from the German Version of the TEL – the “WiwiKom-Test”. The “Wiwikom-Test” was funded from the Federal Ministry of Education and Research (Funding grant number: 01PK11013A). Copyright © 2014 JGU Mainz, FB 03, Wirtschaftspädagogik I, Mainz. All rights reserved. For more information visit http://www.wipaed.uni-mainz.de/ls/1085_ENG_HTML.php. Furthermore, we thank Katharina Schild and Jens Herlemann for their help in coding student answers. Lastly, we thank Taiga Brahm for her comments regarding an earlier version of this manuscript.

1. Introduction

Economists use graphs to present and analyze relationships, most prominently in the supply-and-demand model. Graphs are not only visualizations but also representations of domain principles in the economic domain. Accordingly, the ability to understand and work with graphs is at the core of economic thinking and practicing (Benedict & Hoag, 2011; Davies & Mangan, 2007). Nevertheless, graphs pose certain challenges for novices in the field which might be related to a misunderstanding of central relationships depicted in the graphs but also to misconceptions of how graphs are used in the economic domain (Benedict & Hoag, 2011; Davies & Mangan, 2007; Ring, 2020; Schuhmann et al., 2005; Strober & Cook, 1992). Until recently, researchers in economic education have investigated different research questions regarding the use of graphs in teaching, for example, the effect of adding graphs to lectures in higher education (e.g., Cohn et al., 2001; Cohn & Cohn, 1994; Kourilsky & Wittrock, 1987; Wheat, 2007) and of certain graph formats on conceptual understanding (Chiou, 2009; Jägerskog et al., 2019; Marangos & Alley, 2007; Reingewertz, 2013; Wheat, 2007). Students' knowledge of the relationships depicted in different graphs—such as the rules of supply and demand—has also been part of economics literacy and competence assessment instruments (e.g., Kaiser et al., 2020; Walstad et al., 2013; Walstad & Rebeck, 2008).

Taken together, these research inquiries focus on graphs as learning aids from a teaching perspective and on the underlying mechanisms represented in the graphs. Learners' ability to work with graphs, that is, to perform typical graph operations, has not been a major focus in the economics domain (for exceptions, see Section 2.2). This is important, as assessments that focus on the depicted relationship rather than on the graphs might not reveal learners' misunderstanding of the graph format and its purpose. In other words, learners might know that certain factors influence supply or demand but might still have difficulty understanding that this could be represented as a shift of the one of the curves. To uncover these misconceptions and better understand how learners can be supported in learning economics, it is necessary to analyze learners' ability to work with graphs in the economics domain as procedural knowledge at the intersection of economics competence and generative quantitative abilities.

To address this research gap, we developed a conceptual model with six subcomponents for graph competence in economic education: reading, interpreting, choosing the right graph to display a relationship, using the graph to illustrate a domain-specific problem, identifying errors in a graph, and evaluating the underlying model assumptions. For each of these subcomponents,

we developed items for three types of graphs: the supply and demand model, break-even analysis, and indifference curves.

Using a sample of more than 500 higher education (HE) students with various educational backgrounds, we analyze the quality of the resulting 18-item instrument. We investigate item characteristics with item response theory (IRT) as well as the convergent and criterion validity of the instrument, by examining the relationship between test scores and predictors of economics knowledge and graph competence. Confirming previous research on graph competence, the results show a significant correlation between domain knowledge and graph competence as well as a positive relationship with math grades and previous economics learning opportunities. Item difficulties indicate that questions that go beyond simple graph operations are more challenging for learners in the economics domain.

The results offer practical insights for economics instructors on the use of graphs in teaching. It suggests that learners may not intuitively grasp economic graphs, despite their common use in instruction. The study highlights the importance of discussing not just graph reading rules but also the link between graphs and economic concepts. We provide a conceptual model which can guide the development of diagnostic tests to assess students' graph comprehension and provide a complex and valid instrument that researchers can use to analyze the effect of interventions on graph competence in economic education.

2. Previous research

2.1 Graphs in the economics domain

The main goal of teaching and learning in economics is not only developing factual knowledge but also understanding complex and dynamic relationships that go beyond everyday concepts (Davies & Mangan, 2007). Often, graphs represent these abstract concepts and relationships and therefore can be seen as communication tools of the domain in research (Demir & Tollison, 2015) as well as education (Cohn & Cohn, 1994). A typical example is the relationship among supply, demand, and price in the market model (Mankiw & Taylor, 2020). The graph is not only a visualization – as it allows the reader to quickly understand the relationship – but also an economist's tool for analysis, as each part of the graph represents a variable of interest. Market interventions, such as the minimum price, can be modeled in the graph as a horizontal line, and the effect can be interpreted by analyzing the intersections with the supply-and-demand curves (Mankiw & Taylor, 2020). Consequently, when learners are confronted with graphs, they need to be able to read the graph features (read the point of intersection) and

connect the features to domain principles (e.g., understanding that the intersections represent the effect of the minimum price).

Due to their importance as domain models, graphs have been thoroughly investigated as learning aids from a teaching perspective (Chiou, 2009; Cohn et al., 2001; Cohn & Cohn, 1994; Jägerskog, 2020; Kourilsky & Wittrock, 1987; Marangos & Alley, 2007; Reingewertz, 2013; Wheat, 2007). Cohn et al. (2001), for example, analyzed the learning outcomes of students who were randomly assigned to a lecture with or without graphs and showed that a lecture with graphs was not beneficial or could even lead to adverse learning outcomes. One potential reason, according to the authors, is that the students might not have been able to fully understand the visualized principles or the purpose of the graphs. Consequently, the graphs might have distracted learners rather than supported their learning process. As graphs are difficult for learners, different authors have argued that other visual representations, such as charts of feedback loops, might be less challenging and therefore more effective for teaching (Chiou, 2009; Jägerskog, 2020; Marangos & Alley, 2007; Reingewertz, 2013; Wheat, 2007).

However, graphs are seen as an important communication and analysis tool in the economics domain. Accordingly, the focus should be shifted to *learners' competence* in working with economics graphs. Thus, it is vital to analyze graphs not only as a tool from a teaching perspective but also as a domain-specific competence that learners need to master the economics way of thinking. To identify the necessary abilities and relevant predictors for graph competence, in the following sections, we summarize the literature on economics competence and graph competence in other domains.

2.2 Economics competence and graphs

In research on economic education and economics knowledge and skills, various instruments have been developed for both secondary and university students (e.g., Kaiser et al., 2020; Walstad et al., 2013; Walstad & Rebeck, 2008). These instruments define economic competence as learners' ability to understand and apply economic knowledge from different sub-disciplines such as basic economic principles, markets, the role of government or international economics. The Test of Economic Literacy (TEL) and the Test of Understanding in College Economics (TUCE) are the most prominent instruments in secondary and higher education respectively. As these tests are meant to measure a wide range of economic competencies, learners' ability to work with graphs plays a minor role. Of all items concerning microeconomics on the TUCE (Version 4; Walstad & Rebeck, 2008), for example, one item asks learners to interpret part of a graph. Overall, these tests are not suitable (and were not

meant) for analyzing a learner's graph competence. Nevertheless, as domain knowledge and graph competence are intertwined, some of the predictors of economics competence might also manifest as graph competence in the economics domain.

With the help of these tests, researchers have identified relevant predictors of economics competence, such as gender (in favor of male participants, e.g., Happ et al., 2021) or previous economics learning opportunities (e.g., Brückner et al., 2015). Regarding math-related skills, Schuhmann et al. (2005) showed that general quantitative abilities are necessary for developing economics competence. In their instrument, the authors used items that assess typical graph skills for which participants had to read and interpret data points or identify the slope of a curve with other items, such as solving equations or computing simple percentages as a predictor of general quantitative skills. The authors concluded that students do not do well on these questions and argued that, consequently, they might not understand the language used by instructors in economics courses (Schuhmann et al., 2005). Focusing on graphical analysis, Hill and Stegner (2003) analyzed which students could answer a single graphical question regarding surplus. The authors demonstrated that in addition to students whose mother had at least a bachelor's degree, students with a higher GPA and a preference for math and logic puzzles were more likely to answer the question correctly.

Only a few studies have focused explicitly on the ability of learners to work with graphs in the economics domain (Cohn et al., 2004; Marire, 2017; Ramos Salazar & Hayward, 2022). Using assessment data from 578 students in an introductory microeconomics course, Marire (2017) demonstrated that learners find graph-based multiple-choice questions challenging and are often unable to reproduce graphs in written questions. Notably, the author found differences between the graph content; namely, questions that address the indifference curve were most challenging and further identified a gender bias in favor of male learners (Marire, 2017). The gender gap for graph questions in economics was recently confirmed by Ramos Salazar and Hayward (2022), who found differences between male and female learners in a sample of 206 undergraduate students only for graph questions and not for text questions. The gender differences might partly be explained by attitude toward graphs. In a study by Cohn et al. (2004) female learners reported that graph questions were less helpful for their learning process in economics and more intimidating compared to their male peers.

In summary, in previous instruments, graphs either play a (small) role as content knowledge (Walstad & Rebeck, 2008) or are seen as part of a domain-unspecific quantitative ability (Schuhmann et al., 2005). Little research has yet focused on (measuring) domain-specific graph

understanding (Hill & Stegner, 2003). When graph ability is analyzed in more detail, researchers have shown that graph questions are difficult for all learners and even more so for female students (Cohn et al., 2004; Marire, 2017; Ramos Salazar & Hayward, 2022). To systematically analyze the ability and to investigate specific challenges that arise when students work with graphs, a comprehensive instrument for analyzing graph competence would be helpful. To better understand how graph competence can be modeled and how it is related to domain knowledge, it is fruitful to closely examine other disciplines that have a longer research tradition in investigating graph understanding.

2.3 Measuring graph competence in other domains

Measuring and modeling learners' abilities to work with graphical representations has a significant research tradition in (educational) psychology and math and science education research (e.g., Åberg-Bengtsson & Ottosson, 2006; Beichner, 1994; Berg & Smith, 1994; Curcio, 1987; Friel et al., 2001; Lai et al., 2016; McKenzie & Padilla, 1986; Ring et al., 2019; Shah et al., 2005). We summarize this literature (a) to better understand the relevant skills that learners need when working with graphs, (b) to highlight how they can be meaningfully measured, and (c) to provide an overview of previous predictors of graph competence in other domains.

Regarding the analyzed skills, most frameworks focus on graph reading and comprehension (e.g., Åberg-Bengtsson & Ottosson, 2006; Friel et al., 2001; McKenzie & Padilla, 1986; Ring et al., 2019; Shah et al., 2005). For graph comprehension, according to Friel et al. (2001), it is necessary to have basic math and reading abilities, to know the graph-relevant reading rules and operators, and to have relevant content knowledge to use the representation to answer domain-specific questions. While reading graphs is mostly a generic skill that learners are often able to master, even without domain-specific knowledge, challenges arise when learners need to connect the graph to domain principles in interpretational tasks (Beichner, 1994; Curcio, 1987; Friel et al., 2001; Lachmayer, 2008; Lai et al., 2016). In most domains, learners are not only expected to read and interpret graphs, but should also be able to use them, for example, to illustrate a certain phenomenon. Therefore, more complex competence models include other graph-related abilities, such as graph construction and graph evaluation (Hattikudur et al., 2012; Lachmayer, 2008; Lai et al., 2016). Lai et al. (2016), for example, analyzed the ability of middle school students to interpret, critique and construct different science graphs. The authors showed that learners have difficulty when they try to link graph features to domain concepts. Overall, the results point toward the conclusion that although some learning difficulties might arise from

the graph format, difficulty in understanding graphs is often connected to not understanding the underlying domain principle and how it is represented in the graph.

To meaningfully measure the ability to work with graphs, the item format is a relevant factor. Although multiple-choice items normally lead to higher objectivity, researchers have argued that free response items are necessary to assess certain subcomponents and to adequately capture this complex ability (Berg & Smith, 1994). For example, when the ability of learners to identify errors in a graph is tested, the options of multiple-choice items might prompt learners to discover the mistakes, whereas asking learners to evaluate the graph in an open question might reveal that they are not able to identify mistakes at all. Consequently, newer instruments often combine multiple choice with open-ended items, which often gives deeper insight into the students' ideas regarding graph tasks (Lai et al., 2016).

Several predictors of graph competence have been identified in various studies. One of the most prominent predictors is numerical ability, which is measured either specifically with basic arithmetic operations and number line estimations (e.g., Ludewig et al., 2019) or generally by using math grades in secondary school (Åberg-Bengtsson & Ottosson, 2006; Ring et al., 2019). In some studies, language proficiency and spoken language in the family further play a role as predictors of graph comprehension for secondary students, especially when the focus of the instruments is on graph reading tasks (Åberg-Bengtsson & Ottosson, 2006; Ring et al., 2019). Due to the close link between graph competence and domain-specific knowledge, graph competence test scores usually correlate with domain knowledge test scores, and motivational constructs in the respective domain, for example in biology (Nitz et al., 2014; Ring et al., 2019). Regarding gender differences, studies have shown mixed results. Some have reported a gender gap in favor of male participants (Lowrie & Diezmann, 2011), which is decreased, however, when controlling for mathematical abilities (Åberg-Bengtsson, 1999). Other studies found no differences between male and female learners at all (Lai et al., 2016).

Overall, researchers in other domains have shown that, conceptually, a graph competence model in economic education must encompass different tasks that focus not only on the graph itself but also test learners' ability to connect the graph to the underlying principle. To assess graph understanding, it is important to use different item formats. Important predictors across domains include domain knowledge and interest, as well as mathematical and language abilities.

3. Modeling graph competence in economic education

3.1 Conceptual model

Based on previous research regarding the role of graphs and the assessment of economics and graph competence in different research domains, we developed a conceptual model to measure graph competence in the economics domain. In the first step, we defined the different subcomponents of the typical graph tasks that students need to perform in the economics domain. In the second step, we developed items to fit the different subcomponents. The item formats vary systematically and include complex multiple-choice items (multiple true/false statements), traditional multiple-choice items (one right answer and three distractors), and open response items. The model, as well as the items, was discussed with several experts and economic educators. The resulting model for graph competence comprises six subcomponents, which are summarized in Table 1.

Table 1: Overview of subcomponents of graph competence in economic education

Subcomponent	Description	Item format
Reading	Reading individual components of the graph without content interpretation, e.g., by identifying a certain y-value at a given x-value	Complex multiple choice (multiple true/false statements for one graph)
Interpretation	Domain-specific interpreting graph components, e.g., by identifying the reason for a shift in the supply curve	Complex multiple choice (multiple true/false statements for one graph)
Selection	Selecting a graph component for the representation of a domain-specific relationship, e.g., by choosing the correct curve to represent a good with a low price elasticity of demand	Multiple choice (chose the right graph from four options)
Illustration	Illustrating how the graph can be used to explain a domain principle or an example, e.g., by explaining how the effect of a minimum price could be illustrated	Free response item
Identification of errors	Identifying errors in a given graph, e.g., by noticing mistakes and explaining how they could be corrected	Free response item

Evaluation	Evaluating model assumptions, e.g., by giving reasons why the graph/model cannot be used to explain a certain phenomenon	Free response item
------------	--	--------------------

One of the central challenges when measuring graph understanding within a domain is that graph and content knowledge are conceptually intertwined. Although it is not impossible to answer graph questions without an understanding of economic theory, it is reasonable to assume that learners without prior knowledge have difficulties when more complex graph questions are posed as more complex theoretical knowledge is necessary. Consequently, the instrument measures both the understanding of graphs and of the underlying theory. Compared to instruments that focus solely on the understanding of the theory, we investigate the intersection of graph and content understanding. Some subcomponents (such as reading, interpretation and selection) focus more on the graph itself, whereas others (such as illustration, identifying errors and model evaluation) test whether learners understand how the graph is used within a domain.

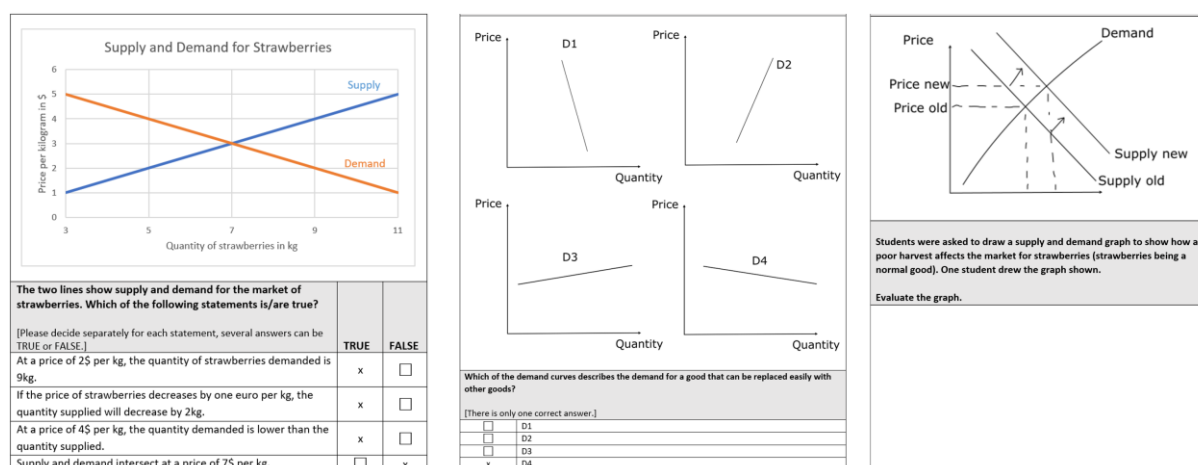
It is generally possible to develop item sets based on the model for different graphs in economic education. For this test instrument, we developed one item per subcomponent for three different graphs: the supply and demand model, break-even analysis, and indifference curves. These graphs were chosen because they cover a range of relevant topics and are included in college textbooks in great detail (for microeconomics: Pindyck & Rubinfeld, 2018; for management and accounting: Drury, 2012). Furthermore, these three graphs represent different forms of functions (equilibrium, linear function, and convex set of curves) and together might provide a glimpse at different difficulties that arise when students work with domain-specific graphs in economics. The first set of items was tested with a small sample of students to ensure comprehensibility and to develop the first draft of the coding manual for the free response items. In the following section, we use three item examples to illustrate subcomponents and testing formats. The final set of items for the supply-and-demand model is listed in Appendix A.

3.2 Example items

To assess learners' ability to read graphs, students had to identify single data points or compute/read off the slope of the graph without any content interpretation. For the supply-and-demand graph, for example, the learners had to check the quantity demanded for strawberries at a certain price level. For selecting, learners had to choose the curve that represented a domain-specific relationship. In the supply-and-demand model, learners had to name the curve, which represented a good that could easily be replaced by other goods. To test learners' ability to

identify mistakes in a given graph, they received a graph that a fictitious student had drawn in response to a certain prompt and were asked to evaluate it. For supply and demand, the graph contained two errors: Supply and demand were interchanged, and the x-axis label was missing. Since we were interested in whether learners noticed the errors at all, we left the instruction unspecified and did not ask them to identify the mistakes directly.

Figure 1: Translated item example for the subcomponents of reading, selecting and identifying mistakes (from left to right)



3.3 Graph item coding procedure

For complex multiple-choice items (multiple true–false statements for one graph), learners could receive partial credit for each correctly marked statement. For example, if a learner correctly marked two out of four statements that the item comprised, they received half the credit. For every free response item, the learners could receive partial credit to the extent that they were able to perform the task. For example, in the item above, in which learners had to identify mistakes in the supply-and-demand graph, they received a third of the credit when they identified only the missing x-axis label, a third of the credit for noticing that the quantity decreased, although it should increase (without realizing that this was the result of the interchanged supply and demand), and full credit when they pointed out that supply and demand were switched. For all open response items, the answers were rated by two expert coders according to a preliminary coding manual that included an expert solution and anchor examples for correct and partially correct answers. In the first coding procedure, the coders marked cases in which they were unable to decide on a final rating. These cases were discussed and thus used to further develop and finalize the coding manual. All answers were rated again with the final coding manual (see Appendix B for the coding manual of the items for the supply-and-demand model). As a measure of interrater reliability, we calculated intraclass correlations (ICC) for every free response item. In the final coding, an average ICC of 0.81 was reached across all

items, which is described as excellent agreement (ICC above 0.75; Cicchetti, 1994). This indicates that the two coders had a high degree of agreement and suggests that the items were rated similarly across the coders. All disagreements after the final coding were solved by discussion among the coders (see Table C1 in Appendix C for auxiliary interrater reliability characteristics).

4. Sample and methodology

4.1 Sample

The test instrument was administered as an online assessment from October 2021 to February 2022 to 579 students at various higher education institutions in southwestern Germany. As this was during the covid pandemic, it was not possible to embed the test in in-person lectures. To ensure that the sample contained students with and without previous knowledge of economics, students in all disciplines were contacted via e-mail and asked to participate in the study. As no specific sampling procedures were used and participation was voluntary, the sample is a convenience sample and is not representative of higher education students in Germany in general.

Table 2 shows the sample characteristics across individual variables. Before the students filled out the graph test, they answered questions regarding basic information, such as gender and age. We also assessed previous economic education in school and their primary fields of study. Self-disclosed math and German grades at the end of secondary education were used as a proxy for general quantitative and language-related skills, respectively. At the end of the test, interest in economics was assessed with four items (e.g., “I like reading about economic topics” adapted from Wilde et al., 2009).

Table 2: Sample characteristics

Variable	N	Mean	SD	Min	Max
Age (years)	574	23.483	5.361	15	69
Female (1/0)	555	69.2 %			
Math grade (0-15)	569	10.155	3.536	0	15
German grade (0-15)	570	10.628	2.681	0	15
Economics at school (1/0)	574	37.5 %			
Semester	568	6.801	4.84	0	28
Teacher training (1/0)	573	21.3 %			
Economics at university (1/0)	579	21.8 %			
Interest in economics	379	3,186	1,365	1	6

Notes: Descriptive statistics are shown for various individual variables. German and math grades are scaled from 0 (worst) to 15 (best). Interest in economics is an equally weighted summary score based on four items (Wilde et al., 2009). The items regarding interest were at the end of the questionnaire; the lower number of learners can be explained by dropouts.

To reduce the overall testing time, the learners answered items regarding only two of the three graphs, which were randomly selected. The questions for one graph were ordered according to the subcomponents described in Table 1. Consequently, every participant was given 12 out of 18 items in a random block-wise order (one participant, for example, first received the items concerning the indifference curve, items 13–18 in Table 3, and then received the items for the supply-and-demand model, items 1–6 in Table 3). Across all participants, the median test taking time for the complete test is 28 minutes, for one block, such as the six items concerning the supply and demand graph, the median is 12 minutes (for cost functions around 11.5 and for indifference curves around 10 minutes). All material and procedures used in the study were reviewed and accepted by the institutional review board of the institution of the first author. All participants were informed about the study and declared their consent for data usage.

4.2 Item response theory

Educational large-scale assessments around the globe (e.g., PISA, TIMSS, and NAEP) have incorporated the use of IRT models to overcome several shortcomings of classical test theory (CTT; see Baker, 2002; Hambleton & Swaminathan, 2013). CTT relies on observed scores (i.e., the sum of correctly solved items); IRT models assume a logistic relationship between the underlying trait (e.g., graph competence) and the probability of endorsing an item. Aside from the widespread use for score-reporting purposes, IRT models allow researchers to investigate several important item characteristics (for applications in the economics domain, see Knoll & Houts, 2012; Ranyard et al., 2020; Walstad & Rebeck, 2017; Walstad & Robson, 1997).

One key requirement for the use of IRT models is that the underlying construct is unidimensional (i.e., our instrument measures only one type of ability). A common method for assessing the dimensional structure is principal component analysis (PCA). In this study, the PCA results produced an eigenvalue of 4.2 for the first component, which accounted for 22.4 percent of the total variance. The second component showed an eigenvalue of only 1.8, with no significant drop in all remaining components. Thus, the first component appears to be dominant, reflecting the unidimensionality of the scale.

To account for the polytomous response format in the item set, we used a partial credit model (PCM; Masters, 1982). Due to the ordered response format (i.e., obtaining two points is more difficult than obtaining zero or one point), item responses reflect the degree of correctness in

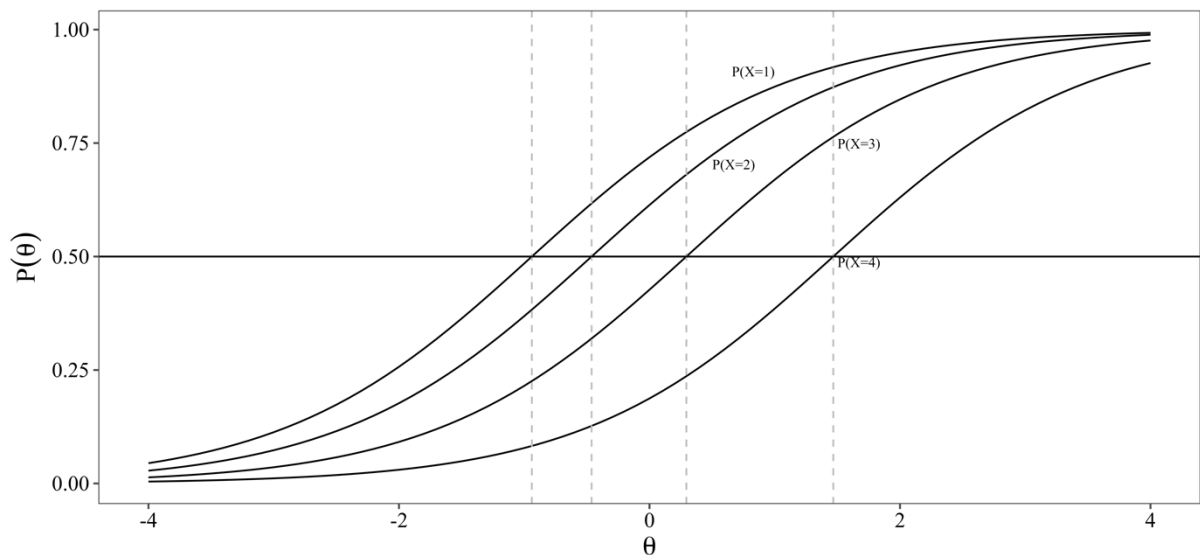
the answer to an item rather than simply being classified as correct or incorrect. The relationship between the probability of obtaining a higher category in the ordinal response format and the underlying latent ability is formally expressed as

$$P(X_i | \theta_v, b_{ih}) = \frac{\exp(\sum_{h=0}^{X_i} (\theta_v - b_{ih}))}{\sum_{h=0}^{X_i} \exp(\sum_{h=0}^{X_i} (\theta_v - b_{ih}h))}, \quad (1)$$

where θ_v denotes the ability of person v and b_{ih} the step difficulty representing the difficulty of obtaining h points over $(h-1)$ points.

The logistic relationship between the probability of achieving the higher adjacent category and the underlying latent trait can be graphically illustrated with operational characteristic curves (OCCs), with the curves for item 8 shown in Figure 1. The threshold parameter (or step difficulty) b reflects the point on the logit scale where $P(X_i = h | \theta_v) = 0.5$ applies, that is, the point on the ability scale where a person v has a greater probability of scoring the item than indicated by chance. For example, item 1 has five response categories; that is, respondents can achieve scores from 0 to 4 following an ordinal structure. Therefore, we expect the threshold parameter (or step difficulty) to be larger for obtaining four points than the parameter for obtaining three points or the next lower adjacent category. Figure 2 illustrates that the response categories for item 1 functioned as expected.

Figure 2: Operational characteristic curves (OCC) for item 1



5. Results

5.1 Item analysis

We investigate item characteristics based on indicators from CTT and the PCM specified in equation (1); the results are shown in Table 3. The corrected item-total correlation (r_{it}) is a widely used statistic in item analysis and measures the (point-biserial) correlation between the item response and the total score on all remaining items. Therefore, the measure reflects the discriminatory power of an item. Higher correlation coefficients suggest that the item discriminates between low- and high-ability learners. The results show that most items had sufficient item-total correlations, with values above the normative threshold of 0.2 (Walstad & Rebeck, 2017). Item 3 appears to be a weaker discriminator, and items 4 and 15 appear to have no discriminatory power.

The IRT model results, in particular the estimated difficulty parameters b_i , indicate that the item set captures a broad range of ability levels, with item 9 the easiest and item 17 the hardest item to solve. We also investigated fit statistics using weighted mean square residuals (Infit) and unweighted fit statistics (Outfit; Wright & Panchapakesan, 1969). Values close to 1 indicate a perfect model fit. The results show that the fit measures are within conventional thresholds ranging from 0.5 to 1.5 (Ames & Penfield, 2015), indicating that the observed responses closely match the responses predicted by the model.

Table 3: Item characteristics

Item no.	N	CTT		IRT				Outfit	Infit
		M	r_{it}	b_1 [SE]	b_2 [SE]	b_3 [SE]	b_4 [SE]		
1	386	1.974	0.499	-0.939 [0.146]	-0.462 [0.195]	0.294 [0.231]	1.467 [0.283]	1.112	1.084
2	386	2.412	0.561	-0.765 [0.151]	-0.481 [0.208]	-0.347 [0.249]	0.457 [0.278]	1.056	1.017
3	304	0.405	0.143	0.65 [0.123]				1.093	1.072
4	238	0.261	0.057	1.441 [0.156]	4.27 [0.734]			1.296	1.102
5	243	0.761	0.327	0.222 [0.139]	1.392 [0.224]	2.924 [0.457]		1.061	1.041
6	210	0.524	0.469	0.65 [0.148]	2.203 [0.288]			0.886	0.899
7	387	2.773	0.531	-0.741 [0.167]	-0.718 [0.236]	-0.625 [0.281]	-0.021 [0.305]	1.002	1.034
8	387	2.09	0.642	-0.726 [0.145]	-0.491 [0.199]	0.064 [0.234]	1.208 [0.276]	0.835	0.863
9	295	0.834	0.269	-1.463 [0.16]				0.886	0.946
10	246	0.756	0.288	-0.883 [0.153]				0.915	0.949
11	260	1.535	0.347	-1.363 [0.218]	-0.15 [0.257]			0.893	0.946
12	242	1.079	0.328	-1.292 [0.186]	1.597 [0.246]			0.963	0.962

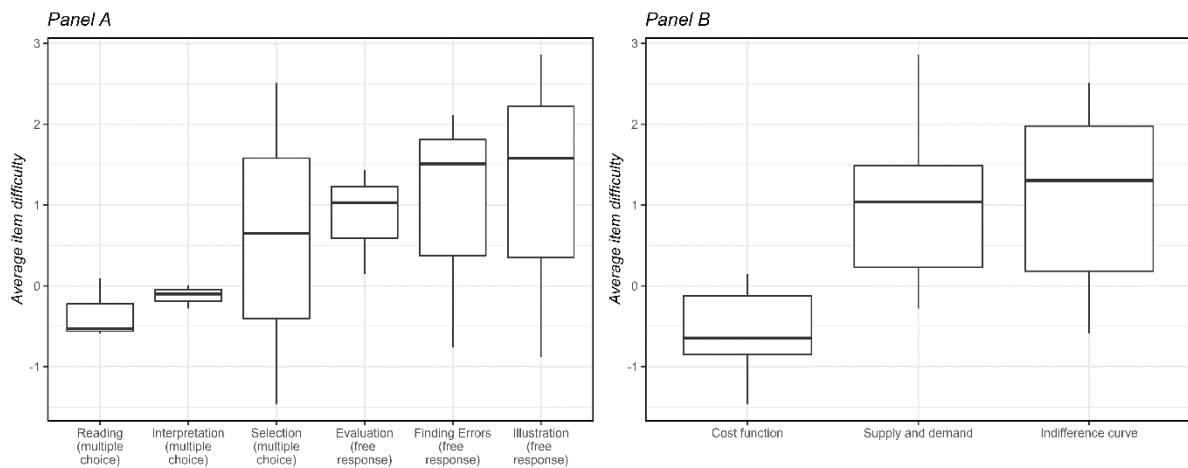
13	387	2.835	0.529	-0.907 [0.175]	-0.851 [0.244]	-0.65 [0.285]	0.063 [0.308]	0.993	1.052
14	387	2.147	0.574	-0.943 [0.149]	-0.436 [0.199]	0.217 [0.236]	0.755 [0.272]	0.985	0.993
15	309	0.11	0.043	2.509 [0.186]				1.209	1.05
16	223	0.466	0.247	0.818 [0.147]	2.333 [0.296]			1.08	1.046
17	226	0.788	0.392	-0.187 [0.147]	2.017 [0.25]	4.507 [1.039]		0.929	0.936
18	231	0.81	0.236	-0.565 [0.152]	2.615 [0.281]			1.01	1.01

Notes: The lower number of learners for the items compared to the overall sample size is a consequence of the test design, which is explained in Section 4.1.

5.2 Item subgroup difficulty

To better understand what drives the difficulty of the different graph tasks, we grouped all items according to the subcomponent of the competence (e.g., reading), as well as the graph content, and compared the difficulty estimated from the IRT model. As the test consists of one item per subcomponent per graph, we display the range of difficulty of three items per subcomponent (Figure 3, panel A) and the range of difficulty of six items per graph (Figure 3, panel B). The figure shows that, on average, multiple-choice items that require reading and interpretation are easier than free response items, which involve more complex operations, such as using the graph to illustrate an example or to identify errors. Furthermore, items involving the cost function are less difficult than items involving the supply-and-demand model or indifference curves.

Figure 3: Distribution of difficulty per subcomponent (panel A) and per graph (panel B)



5.3 Convergent validity

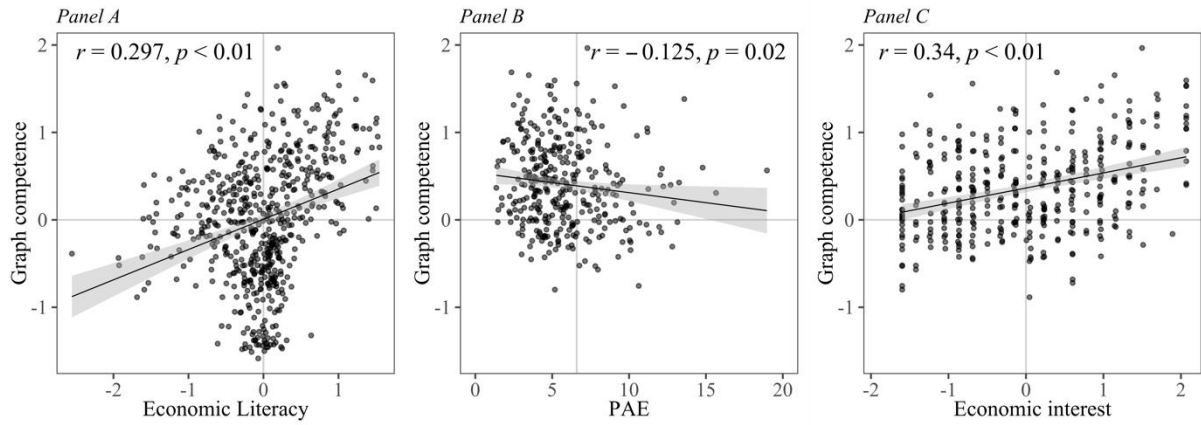
Evidence of the construct validity of the instrument stems from correlations with adjacent constructs. For that purpose, we estimate individual graph competence scores using the IRT model specified in equation (1). As estimated ability scores are generally subject to

measurement errors, we implement a multiple imputation approach to correct for the error term based on a latent regression model (plausible values; Marsman et al., 2016). Specifically, we estimate 20 different plausible values and aggregate them iteratively in the subsequent correlation analysis.

To analyze the convergent validity of the newly developed measure, different scales for well-established predictors of economics and graph competence were included in the questionnaire. For content-oriented knowledge in the economics domain, a short scale of the Test of Economic Literacy (Walstad et al., 2001) was used in the version translated and validated for the German context by Happ et al. (2018).¹ As with the graph competence scores, we rely on 20 plausible values extracted from a unidimensional one-parameter IRT model for dichotomous response formats (Rasch, 1960). In addition, number line estimation items, in which the learners were asked to estimate a given number on a number line, serve as a proxy for understanding numerical quantities and their measurement. The difference between the correct position and the estimated position was expressed in terms of the percentage of the number line length; the percentage of absolute error (PAE) was used as a variable. As the measure indicates the deviation of the estimate from the correct position, higher PAE values indicate lower ability. The correlation analysis results (Figure 4, panel A) reveal that the scores on the TEL are significantly associated with graph competence scores ($r = 0.297$; $p < 0.01$), indicating that the items measuring graph competence capture the domain-specific dimension. Furthermore, the negative correlation between the graph competence scores and the PAE in panel B indicates that the instrument captures differences in the generic understanding of numerical quantities. Finally, interest in economics may serve as a prerequisite for acquiring domain-specific competences. Panel C in Figure 4 shows that interest in economics is positively associated with graph competence scores ($r = 0.34$; $p < 0.01$; panel C).

Figure 4: Correlations with adjacent constructs

¹ We thank the university of Mainz and namely Prof. Dr. Zlatkin-Troitschanskaia and Prof. Dr. Roland Happ for the opportunity to use items from the German Version of the TEL – the “WiwiKom-Test”.



5.4 Differential item functioning (DIF)

Analyzing differential item functioning (DIF) provides additional evidence of the construct validity and test fairness of the instrument (Holland & Wainer, 2012). An item is flagged as DIF if subgroups with the same latent ability show different probabilities of solving an item correctly or reaching the next category when the item is polytomously scored. Logistic regression is one of the most popular frameworks for detecting DIF because of its capability of identifying uniform and nonuniform DIF (Swaminathan & Rogers, 1990). When uniform DIF occurs, the probability of endorsing an item (or reaching the next category) is constantly higher for one subgroup, whereas nonuniform DIF indicates that the difference in probability may vary along the latent trait continuum.

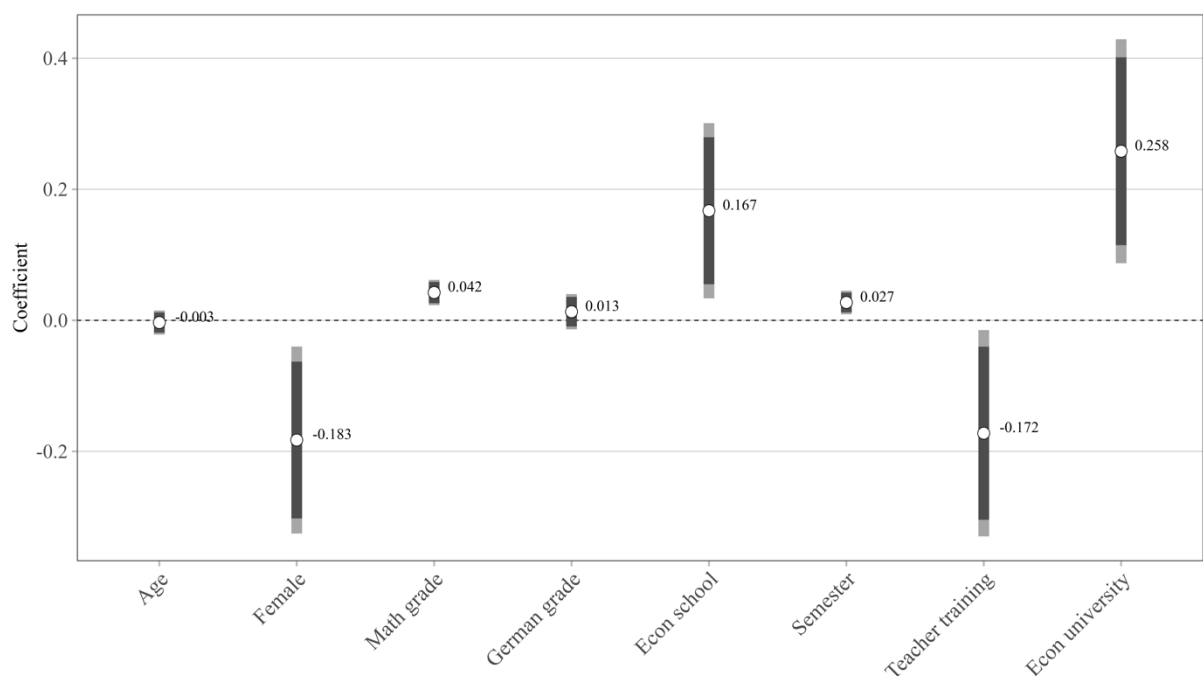
To detect DIF in polytomously scored items, we use the ordinal logistic regression approach described by Zumbo (1999). We regress latent trait estimates θ , the group membership (e.g., being female), and the interaction between these two components on the log odds of reaching the higher category (y), formally expressed as $y = \beta_0 + \beta_1\theta + \beta_2\text{Group} + \beta_3(\theta \times \text{Group}) + \epsilon$. To categorize the severity of DIF, we follow the scheme provided by Zumbo and Thomas (1997), in which items are flagged as negligible nonuniform DIF (category A) if the difference in the pseudo- R^2 between the aforementioned regression model and the reduced model without an interaction term is less than 0.13 ($\Delta R^2 \leq 0.13$). The results for gender, being enrolled in teacher training, and whether learners had received economic education in school are reported in Table C2 in Appendix C. Overall, all items exhibit pseudo- R^2 below 0.13 across all three demographic criteria, indicating no or only negligible uniform and nonuniform DIF.

5.4 Criterion validity

In this section, we investigate the correlations between test scores and individual variables, which serve as suggestive evidence of the instrument's criterion validity. Specifically, we estimate the linear regression model (OLS) $y_v = \beta_0 + \sum \beta_j X_j + \epsilon_v$, where y denotes the IRT

score from 20 multiple imputations (plausible values), and X_j is a covariate vector with the variables listed in Table 2. ε_v represents the idiosyncratic error term. The data reveal that the test scores were positively associated with the learners' math proficiency. A one-unit increase in a math grade is associated with an increase in the test score by 0.04 standard deviation units. In contrast, learners' German grades do not correlate with test scores, which serves as suggestive evidence for the discriminant validity of the scale. The instrument was created to measure graph competence in the economics domain. Therefore, we expected that learners with previous economic education would achieve higher test scores relative to learners with no economic education. As shown in Figure 5, the results support this expectation. Learners who had economic education in secondary school (Econ school) or university (Econ university) scored significantly higher by a magnitude of 0.17 and 0.26 standard deviation units, respectively. This indicates that the domain-specific part of the underlying construct was captured with the test instrument. Regarding gender, we observe mean differences in favor of male students in the expected direction (Cohn et al., 2004; Marire, 2017; Ramos Salazar & Hayward, 2022). Collectively, the analyses show correlations that mirror the estimates documented in previous work.

Figure 5: Regression estimates (OLS)



Notes: Coefficients from the regression model (OLS) with 95% CIs (dark gray) and 90% CIs (light gray) at an estimated intercept of 0.768. The dependent variable *graph competence* is imputed and pooled from latent regression (plausible values). Non-categorical variables are mean centered. The number of observations is $N = 548$.

6. Discussion

In this study, we analyzed the properties of a test instrument for assessing graph competence in economic education. The instrument was based on a new conceptual model that included six subcomponents: reading, interpreting, selecting, illustrating, finding errors, and evaluating the model. For each subcomponent, one item was developed for typical economics graphs: the supply-and-demand model, cost function, and indifference curves. Analyzing data from 579 tertiary students in different disciplines with IRT, we showed that the test instrument is suitable for measuring the underlying ability of graph competence and that the items capture a broad range of ability levels.

In line with previous studies in economic education (Cohn & Cohn, 1994; Cohn et al., 2001, 2004; Kourilsky & Wittrock, 1987; Marire, 2017; Wheat, 2007), the present study results confirm that graphs are not intuitively understood by learners. In particular, this study showed that more elaborate graph tasks are difficult to master. In other words, although learners might be able to read and interpret surface features, free response tasks that require a deeper understanding of the graph as a tool in the domain are more challenging. Furthermore, questions regarding the cost function are easier compared to the supply-and-demand model and indifference curves, which partly replicates Marire's (2017) findings. A possible explanation might be that linear functions are more familiar to the diverse set of learners in this sample than more domain-specific equilibrium models and convex functions.

One noteworthy result of this study is the gender gap in favor of male learners. Although we cannot explain the differences, this result is interesting because it highlights the persistence of performance differences between male and female learners in economics knowledge, specifically for graphs in economic education (Cohn et al., 2004; Marire, 2017; Ramos Salazar & Hayward, 2022). As gender differences are not always visible in studies that analyze graph skills in other domains (Lai et al., 2016), it remains an open question whether the lower performance of female learners on graph tasks is due to the graph domain rather than the graph format. Following Cohn et al. (2004), a potential approach could be to analyze gender differences in attitudes regarding graphs and their helpfulness in learning economics. In particular, because graphs are widely used in introductory classes, more research on the challenges with graphs (for certain subgroups) is warranted.

With this study, we also replicated previous findings regarding the relationship between math skills and graph competence (Hill & Stegner, 2003; Schuhmann et al., 2005), although mathematics does not seem to be the strongest predictor in this study. A likely explanation is

the use of self-disclosed mathematics grades at the end of secondary school as an indicator of students' math skills, which is a very general measure and (specifically for advanced students) might not be a good estimation of students' current math skills. To better understand the relationship between domain-specific graph competence and math abilities, future research could further investigate this relationship by using more specific measures for domain-relevant math skills (e.g., Orlov et al., 2021).

Since the economic graph test was meant to measure the abilities at the intersection of graph and theoretical understanding, we assumed this conceptual link to be visible in our study. As expected, graph competence is positively associated with economics knowledge, interest in economics, and previous economics learning opportunities. The low correlation coefficient suggests that although economic literacy and graph competence are connected, they are not identical skills. Of course, we cannot draw causal conclusions from the relationships, and it remains unclear whether graph competence is the result (or a prerequisite) of economics competence, interest, and education. To better understand the impact of different learning opportunities, future research could use the test instrument in a longitudinal study to investigate the competence development of students who receive economic education.

On a methodical level, these results support the argument originally made by Berg and Smith (1994): Free response items might be necessary to adequately test graph competence in the economics domain. One methodical challenge in that regard is the joint nature of task complexity and item format. Future research could partly disentangle this problem by using only free response items to compare subcomponent difficulty. Using only multiple-choice items would produce higher objectivity but might lead to lower validity, as they test whether learners can choose the correct but available solution rather than using the graph to develop a solution on their own.

The findings of this research provide practical insights for instructors regarding the role of graphs in teaching and learning economics. Our work suggests that teachers should not assume that economic graphs are intuitively understood by learners. This is highly relevant because graphs are often used by instructors as a means of promoting understanding of an underlying economic concept (such as price formation for the supply and demand model) without necessarily devoting much time to an explanation of its usage. Specifically, instructors should be aware that although general graph reading rules (such as reading single points) do not seem to be a large issue, understanding the relationship between graph(feature) and domain concept is difficult for students, more so for learners without prior economic education. Our work – and

the larger literature body of graph research – indicate that spending class time to explicitly discuss not only graph reading rules but also how the graph is used in the specific context might help students to better grasp economic concepts in the future. Although the provided instrument might be too complex to be integrated regularly into teaching as a diagnostic tool, the instrument, as well as the conceptual model, can guide instructors in developing graph-specific tests to better understand their students' challenges.

To evaluate the implications and generalizability of the results, it is necessary to discuss the circumstances regarding the sample and the test administration. The test was used with a convenience sample of tertiary students in southwestern Germany in a low-stakes online setting. Since more women than men took part in the test and the participation was completely voluntary, we cannot rule out that the differences between male and female students might be a result of a third variable which influences the decision to partake in the test. Furthermore, as the online setting allows little control of learners' test-taking behavior, it might also be worthwhile to administer the test in a more controlled context (e.g., a pen-and-paper test in a lecture hall) to obtain a more precise estimate of students' abilities. Since we were interested in the relationship between economic learning opportunities and graph competence, our sample consisted of students with and without a background in economics. Although this gives a broad estimate of the ability that learners might have developed as a result of their school education, further refining the instrument to investigate economics students' understanding will provide more specific guidance on how to address challenges that arise from the graph format.

The implications of this study are further limited by the selection of certain economics graphs and the limited number of items for certain subcomponents. In that regard, the difference between the item difficulties between subcomponents and graph content should be interpreted with caution. Future studies could use the underlying conceptual model as a blueprint to develop more items for specific subcomponents or use the model with other graphs in the economics domain. In the conceptual model and consequently, in the test instrument, we focused on tasks in which learners had to work with existing graphs rather than construct graphs themselves. From our point of view, investigating students' graph construction abilities could be an interesting avenue for future research to advance the understanding of the challenges that students face in learning economics.

Taken together, the results replicate previous results in graph and economic education research (Cohn & Cohn, 1994; Cohn et al., 2001, 2004; Kourilsky & Wittrock, 1987; Marire, 2017; Wheat, 2007) and provide evidence for the validity of the presented test instrument. Overall,

we contribute a conceptual model that defines different subcomponents of graphing ability in economics and provides a solid instrument that can be used in future research to measure and analyze graph competence in economic education. We identified several predictors and showed the importance of investigating graph competence in greater detail. In addition, the empirical findings in this study provide new insights into the specific challenges that learners face with different domain graphs and, therefore, contribute to our understanding of teaching and learning economics.

References

- Åberg-Bengtsson, L. (1999). Dimensions of performance in the interpretation of diagrams, tables, and maps: Some gender differences in the Swedish scholastic aptitude test. *Journal of Research in Science Teaching*, 36(5), 565–582.
[https://doi.org/10.1002/\(SICI\)1098-2736\(199905\)36:5<565::AID-TEA4>3.0.CO;2-L](https://doi.org/10.1002/(SICI)1098-2736(199905)36:5<565::AID-TEA4>3.0.CO;2-L)
- Åberg-Bengtsson, L., & Ottosson, T. (2006). What lies behind graphicacy? Relating students' results on a test of graphically represented quantitative information to formal academic achievement. *Journal of Research in Science Teaching*, 43(1), 43–62.
<https://doi.org/10.1002/tea.20087>
- Ames, A. J., & Penfield, R. D. (2015). An NCME Instructional Module on Item-Fit Statistics for Item Response Theory Models. *Educational Measurement: Issues and Practice*, 34(3), 39–48. <https://doi.org/10.1111/emip.12067>
- Baker, F. B. (2002). *The basics of item response theory* (2nd ed.). ERIC Clearinghouse on Assessment and Evaluation.
- Beichner, R. J. (1994). Testing student interpretation of kinematics graphs. *American Journal of Physics*, 62(8), 750–762. <https://doi.org/10.1119/1.17449>
- Benedict, M. E., & Hoag, J. (2011). Factors influencing performance in economics: Graphs and quantitative usage. In G. M. Hoyt & K. McGoldrick (Eds.), *International handbook on teaching and learning economics* (pp. 334–340). Edward Elgar Publishing.
- Berg, C. A., & Smith, P. (1994). Assessing students' abilities to construct and interpret line graphs: Disparities between multiple-choice and free-response instruments. *Science Education*, 78(6), 527–554. <https://doi.org/10.1002/sce.3730780602>
- Brückner, S., Förster, M., Zlatkin-Troitschanskaia, O., & Walstad, W. B. (2015). Effects of prior economic education, native language, and gender on economic knowledge of first-year students in higher education. A comparative study between Germany and the USA. *Studies in Higher Education*, 40(3), 437–453.
<https://doi.org/10.1080/03075079.2015.1004235>
- Chiou, C.-C. (2009). Effects of concept mapping strategy on learning performance in business and economics statistics. *Teaching in Higher Education*, 14(1), 55–69.
<https://doi.org/10.1080/13562510802602582>
- Cicchetti, D. V. (1994). Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instruments in Psychology. *Psychological Assessment*, 6(4), 284–290. <https://doi.org/10.1037/1040-3590.6.4.284>

- Cohn, E., & Cohn, S. (1994). Graphs and learning in principles of economics. *The American Economic Review*, 84(2), 197–200.
- Cohn, E., Cohn, S., Balch, D. C., & Bradley, J. (2001). Do Graphs Promote Learning in Principles of Economics? *The Journal of Economic Education*, 32(4), 299–310. <https://doi.org/10.1080/00220480109596110>
- Cohn, E., Cohn, S., Balch, D. C., & Bradley, J. (2004). The Relation between Student Attitudes toward Graphs and Performance in Economics. *The American Economist*, 48(2), 41–52. <https://doi.org/10.1177/056943450404800203>
- Curcio, F. R. (1987). Comprehension of Mathematical Relationships Expressed in Graphs. *Journal for Research in Mathematics Education*, 18(5), 382–393. <https://doi.org/10.2307/749086>
- Davies, P., & Mangan, J. (2007). Threshold concepts and the integration of understanding in economics. *Studies in Higher Education*, 32(6), 711–726. <https://doi.org/10.1080/03075070701685148>
- Demir, I., & Tollison, R. D. (2015). Graphs in Economics. *Economics Bulletin*, 35(3), 1834–1847.
- Drury, C. (2012). *Management and cost accounting* (8. ed.). Cengage Learning.
- Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making Sense of Graphs: Critical Factors Influencing Comprehension and Instructional Implications. *Journal for Research in Mathematics Education*, 32(2), 124–158. <https://doi.org/10.2307/749671>
- Hambleton, R. K., & Swaminathan, H. (2013). *Item Response Theory: Principles and Applications*. Springer Science & Business Media.
- Happ, R., Kato, M., & Rüter, I. (2021). Results from the test of economic literacy in Germany and Japan: A critical discussion on the gender effect. *Citizenship, Social and Economics Education*, 20(1), 48–68. <https://doi.org/10.1177/20471734211004117>
- Happ, R., Zlatkin-Troitschanskaia, O., & Förster, M. (2018). How prior economic education influences beginning university students' knowledge of economics. *Empirical Research in Vocational Education and Training*, 10(1). <https://doi.org/10.1186/s40461-018-0066-7>
- Hattikudur, S., Prather, R. W., Asquith, P., Alibali, M. W., Knuth, E. J., & Nathan, M. (2012). Constructing Graphical Representations: Middle Schoolers' Intuitions and Developing Knowledge About Slope and Y-intercept. *School Science and Mathematics*, 112(4), 230–240. <https://doi.org/10.1111/j.1949-8594.2012.00138.x>

- Hill, C. D., & Stegner, T. (2003). Which Students Benefit from Graphs in a Principles of Economics Class? *The American Economist*, 47(2), 69–77.
<http://www.jstor.org/stable/25604281>
- Holland, P. W., & Wainer, H. (2012). *Differential Item Functioning*. Taylor and Francis.
<https://doi.org/10.4324/9780203357811>
- Jägerskog, A.-S. (2020). *Making Possible by Making Visible: Learning through Visual Representations in Social Science [Doctoral Dissertation, Stockholm University]*. Department of Humanities and Social Sciences Education, Stockholm University.
- Jägerskog, A.-S., Davies, P., & Lundholm, C. (2019). Students' understanding of causation in pricing: A phenomenographic analysis. *Journal of Social Science Education (JSSE)*, 18(3), 89–107. <https://doi.org/10.4119/JSSE-1421>
- Kaiser, T., Oberrauch, L., & Seeber, G. (2020). Measuring economic competence of secondary school students in Germany. *The Journal of Economic Education*, 51(3-4), 227–242. <https://doi.org/10.1080/00220485.2020.1804504>
- Knoll, M. A. Z., & Houts, C. R. (2012). The Financial Knowledge Scale: An Application of Item Response Theory to the Assessment of Financial Literacy. *Journal of Consumer Affairs*, 46(3), 381–410. <https://doi.org/10.1111/j.1745-6606.2012.01241.x>
- Kourilsky, M., & Wittrock, M. C. (1987). Verbal and graphical strategies in the teaching of economics. *Teaching and Teacher Education*, 3(1), 1–12.
[https://doi.org/10.1016/0742-051X\(87\)90030-8](https://doi.org/10.1016/0742-051X(87)90030-8)
- Lachmayer, S. (2008). *Entwicklung und Überprüfung eines Strukturmodells der Diagrammkompetenz für den Biologieunterricht [Development and validation of a graph competence model for biology education] (Doctoral dissertation, Christian-Albrechts-Universität zu Kiel)*. <https://d-nb.info/1019667389/34>
- Lai, K., Cabrera, J., Vitale, J. M., Madhok, J., Tinker, R., & Linn, M. C. (2016). Measuring Graph Comprehension, Critique, and Construction in Science. *Journal of Science Education and Technology*, 25(4), 665–681. <https://doi.org/10.1007/s10956-016-9621-9>
- Lowrie, T., & Diezmann, C. M. (2011). Solving graphics tasks: Gender differences in middle-school students. *Learning and Instruction*, 21(1), 109–125.
<https://doi.org/10.1016/j.learninstruc.2009.11.005>
- Ludewig, U., Lambert, K., Dackermann, T., Scheiter, K., & Möller, K. (2019). Influences of basic numerical abilities on graph reading performance. *Psychological Research*, 1–13. <https://doi.org/10.1007/s00426-019-01144-y>

- Mankiw, N. G., & Taylor, M. P. (2020). *Economics* (Fifth edition). Cengage.
- Marangos, J., & Alley, S. (2007). Effectiveness of concept maps in economics: Evidence from Australia and USA. *Learning and Individual Differences*, 17(2), 193–199.
<https://doi.org/10.1016/j.lindif.2007.03.003>
- Marire, J. (2017). Does a graph-intensive economics curriculum promote epistemological access to economic theory? *South African Journal of Higher Education*, 32(1).
<https://doi.org/10.20853/32-1-1651>
- Marsman, M., Maris, G., Bechger, T., & Glas, C. (2016). What can we learn from Plausible Values? *Psychometrika*, 81(2), 274–289. <https://doi.org/10.1007/s11336-016-9497-x>
- Masters, G. N. (1982). A rasch model for partial credit scoring. *Psychometrika*, 47(2), 149–174. <https://doi.org/10.1007/BF02296272>
- McKenzie, D. L., & Padilla, M. J. (1986). The construction and validation of the test of graphing in science (togs). *Journal of Research in Science Teaching*, 23(7), 571–579.
<https://doi.org/10.1002/tea.3660230702>
- Nitz, S., Ainsworth, S. E., Nerdel, C., & Prechtl, H. (2014). Do student perceptions of teaching predict the development of representational competence and biological knowledge? *Learning and Instruction*, 31, 13–22.
<https://doi.org/10.1016/j.learninstruc.2013.12.003>
- Orlov, G., McKee, D., Foster, I. R., Botton, D., & Thomas, S. R. (2021). Identifying Students at Risk Using a New Math Skills Assessment. *AEA Papers and Proceedings*, 111, 97–101. <https://doi.org/10.1257/pandp.20211044>
- Pindyck, R. S., & Rubinfeld, D. L. (2018). *Microeconomics* (Ninth edition). *The Pearson series in economics*. Pearson.
- Ramos Salazar, L., & Hayward, S. L. (2022). Exploring the Associations of the Bicycle Drawing Test (BDT) on Text-Based and Graphical Assessments in Undergraduate-Level Economics Courses. *College Teaching*, 70(3), 337–349.
<https://doi.org/10.1080/87567555.2021.1944040>
- Ranyard, R., McNair, S., Nicolini, G., & Duxbury, D. (2020). An item response theory approach to constructing and evaluating brief and in-depth financial literacy scales. *Journal of Consumer Affairs*, 54(3), 1121–1156. <https://doi.org/10.1111/joca.12322>
- Rasch, G. (1960). *Probabilistic model for some intelligence and achievement tests*. Danish Institute for Educational Research.
- Reingewertz, Y. (2013). Teaching macroeconomics through flowcharts. *International Review of Economics Education*, 14, 86–93. <https://doi.org/10.1016/j.iree.2013.10.004>

- Ring, M. (2020). *Visual Representations in Economic Education From an Interdisciplinary Perspective [Doctoral Dissertation, Tübingen University]*.
<http://dx.doi.org/10.15496/publikation-51233>
- Ring, M., Brahm, T., & Randler, C. (2019). Do difficulty levels matter for graphical literacy? A performance assessment study with authentic graphs. *International Journal of Science Education*, 41(13), 1787–1804.
<https://doi.org/10.1080/09500693.2019.1640915>
- Schuhmann, P. W., McGoldrick, K [KimMarie], & Burrus, R. T. (2005). Student Quantitative Literacy: Importance, Measurement, and Correlation with Economic Literacy. *The American Economist*, 49(1), 49–65. <https://doi.org/10.1177/056943450504900104>
- Shah, P., Freedman, E. G., & Vekiri, I. (2005). The Comprehension of Quantitative Information in Graphical Displays. In P. Shah & A. Miyake (Eds.), *Cambridge Handbooks in Psychology. The Cambridge Handbook of Visuospatial Thinking* (pp. 426–476). Cambridge University Press.
<https://doi.org/10.1017/CBO9780511610448.012>
- Strober, M. H., & Cook, A. (1992). Economics, lies, and videotapes. *The Journal of Economic Education*, 23(2), 125–151.
- Swaminathan, H., & Rogers, H. J. (1990). Detecting Differential Item Functioning Using Logistic Regression Procedures. *Journal of Educational Measurement*, 27(4), 361–370. <https://doi.org/10.1111/j.1745-3984.1990.tb00754.x>
- Walstad, W. B., & Rebeck, K. (2008). The Test of Understanding of College Economics. *The American Economic Review*, 98(2), 547–551. <https://doi.org/10.1257/aer.98.2.547>
- Walstad, W. B., & Rebeck, K. (2017). The Test of Financial Literacy : Development and measurement characteristics. *The Journal of Economic Education*, 48(2), 113–122.
<https://doi.org/10.1080/00220485.2017.1285739>
- Walstad, W. B., Rebeck, K., & Butters, R. B. (2001). *Test of economic literacy*. National Council on Economic Education.
- Walstad, W. B., Rebeck, K., & Butters, R. B. (2013). The Test of Economic Literacy: Development and Results. *The Journal of Economic Education*, 44(3), 298–309.
<https://doi.org/10.1080/00220485.2013.795462>
- Walstad, W. B., & Robson, D. (1997). Differential Item Functioning and Male-Female Differences on Multiple-Choice Tests in Economics. *The Journal of Economic Education*, 28(2), 155–171. <https://doi.org/10.1080/00220489709595917>

- Wheat, D. (2007). The feedback method of teaching macroeconomics: is it effective? *System Dynamics Review*, 23(4), 391–413.
- Wilde, M., Bätz, K., Kovaleva, A., & Urhahne, D. (2009). Überprüfung einer Kurzsкала intrinsischer Motivation (KIM): [Testing a short scale of intrinsic motivation]. *Zeitschrift Für Didaktik Der Naturwissenschaften*, 15, 31–45.
- Wright, B., & Panchapakesan, N. (1969). A Procedure for Sample-Free Item Analysis. *Educational and Psychological Measurement*, 29(1), 23–48.
<https://doi.org/10.1177/001316446902900102>
- Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-Type (Ordinal) Item Scores*. Directorate of Human Resources Research and Evaluation, Department of National Defense.
<https://faculty.educ.ubc.ca/zumbo/dif/handbook.pdf>
- Zumbo, B. D., & Thomas, D. R. (1997). *A measure of effect size for a model-based approach for studying DIF: Prince George, Canada: University of Northern British Columbia, Edgeworth Laboratory for Quantitative Behavioral Science*.

Supplementary material

to accompany

Measuring Economic Graph Competence

Appendix A: Example Items

Appendix B: Coding manual

Appendix C: Auxiliary results

Appendix A: Items

Graph Competence in Economic Education Item set

(Translated from the original German)

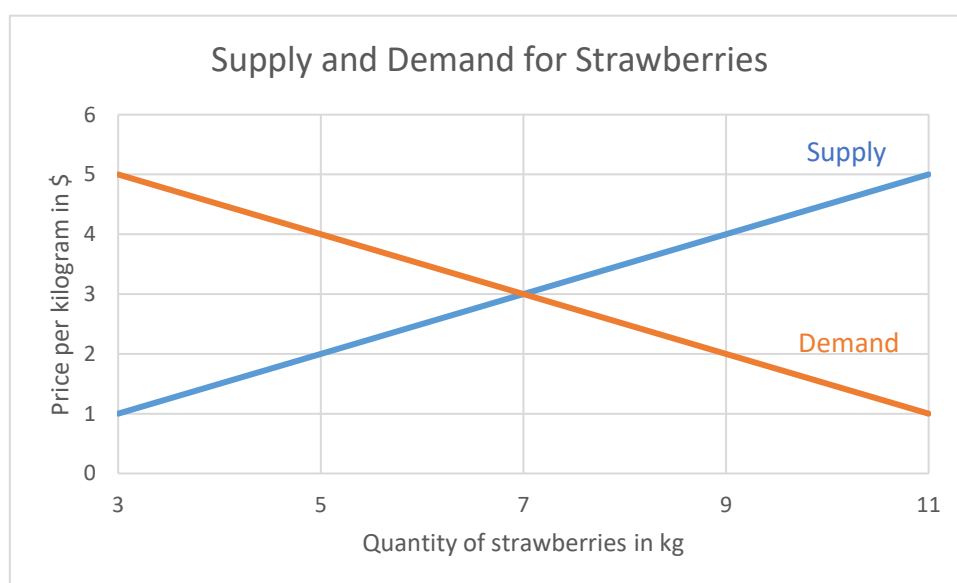
Item Info

Short: DP01

Graph: Supply and Demand

Subcomponent: Reading

Answer options randomized: yes



The two lines show supply and demand for the market of strawberries. Which of the following statements is/are true?

[Please decide separately for each statement, several answers can be TRUE or FALSE.]

	TRUE	FALSE
At a price of 2\$ per kg, the quantity of strawberries demanded is 9kg.	<input type="checkbox"/>	<input type="checkbox"/>
If the price of strawberries decreases by one euro per kg, the quantity supplied will decrease by 2kg.	<input type="checkbox"/>	<input type="checkbox"/>
At a price of 4\$ per kg, the quantity demanded is lower than the quantity supplied.	<input type="checkbox"/>	<input type="checkbox"/>
Supply and demand intersect at a price of 7\$ per kg.	<input type="checkbox"/>	<input type="checkbox"/>

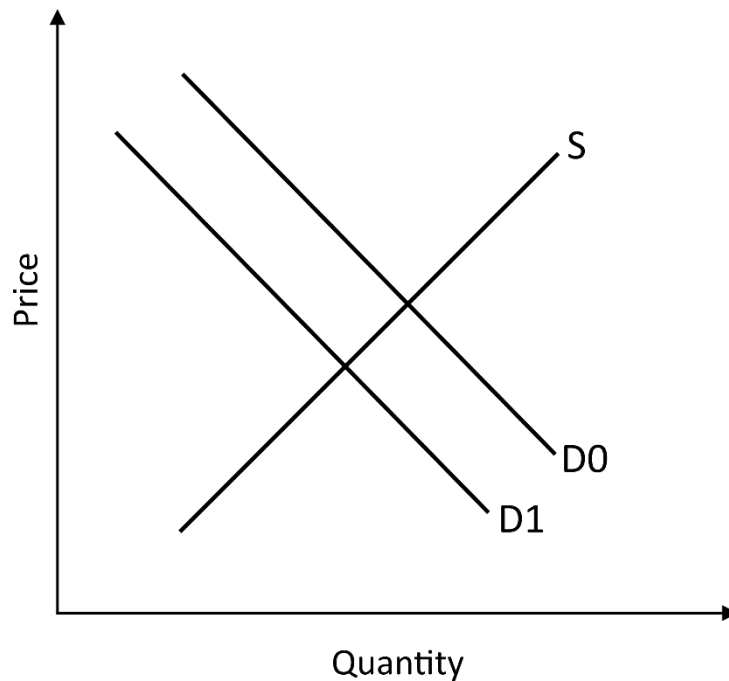
Item Info

Short: DP02

Graph: Supply and Demand

Subcomponent: Interpretation

Answer options randomized: yes



The two straight lines S and D0 show the initial supply and demand of a new smartphone. Which of the following events could lead to the shift of the demand curve from D0 to D1?

[Please decide separately for each statement, several answers can be TRUE or FALSE.]

The smartphone producer announces technical problems of the smartphone in a press conference.

TRUE**FALSE**
☐
☐

Cargo ships are unable to deliver important inputs for producing the smartphone.

☐
☐

The manufacturer uses more efficient production strategies.

☐
☐

The company has completed a successful advertising campaign for the smartphone.

☐
☐

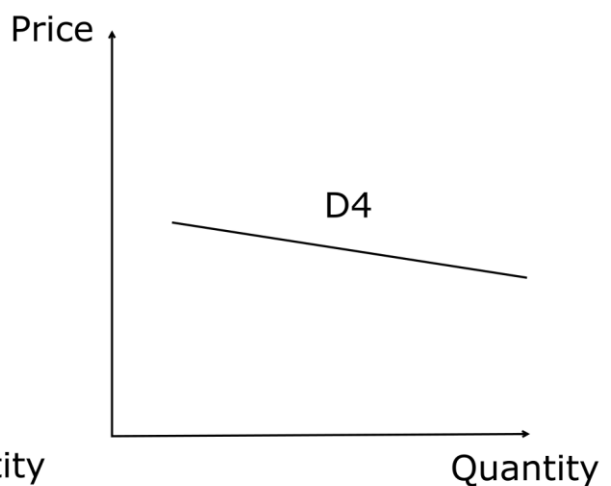
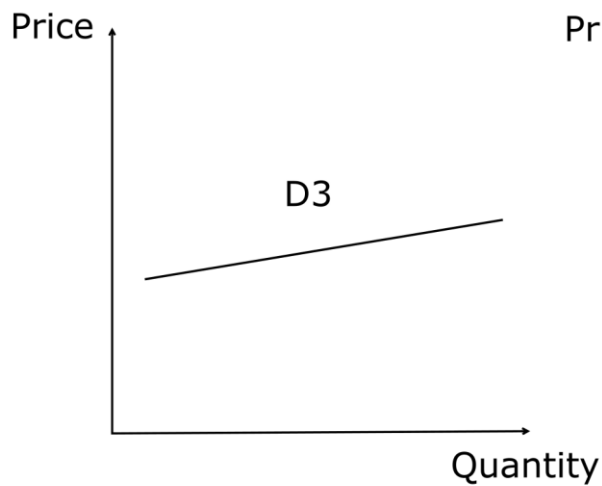
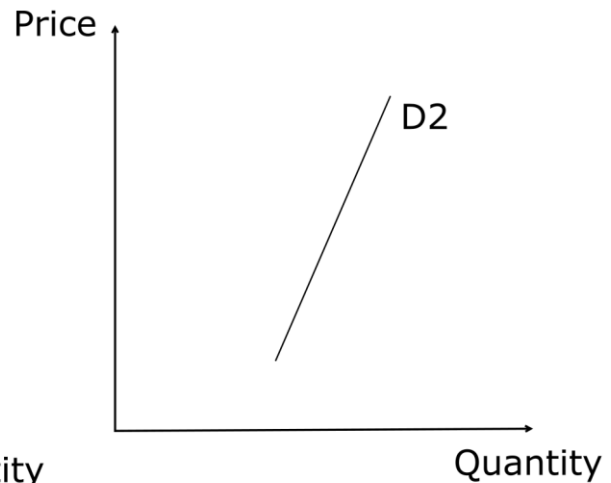
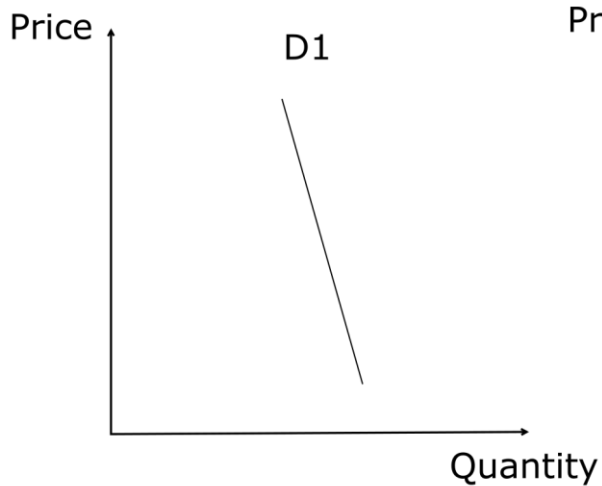
Item Info

Short: DP03

Graph: Supply and Demand

Subcomponent: Graph Selection

Answer options randomized: no



Which of the demand curves describes the demand for a good that can be replaced easily with other goods?

[There is only one correct answer.]

<input type="checkbox"/>	D1
<input type="checkbox"/>	D2
<input type="checkbox"/>	D3
<input type="checkbox"/>	D4

Item Info

Short: DP04

Graph: Supply and Demand

Subcomponent: Illustration

Explain how the supply and demand graph can be used to outline the effect of a binding minimum price (a price above the market clearing price) on the quantity supplied and demanded.

[For precise coding and anchor examples -> see coding manual]

[1 point for minimum price as a horizontal line in the graph]

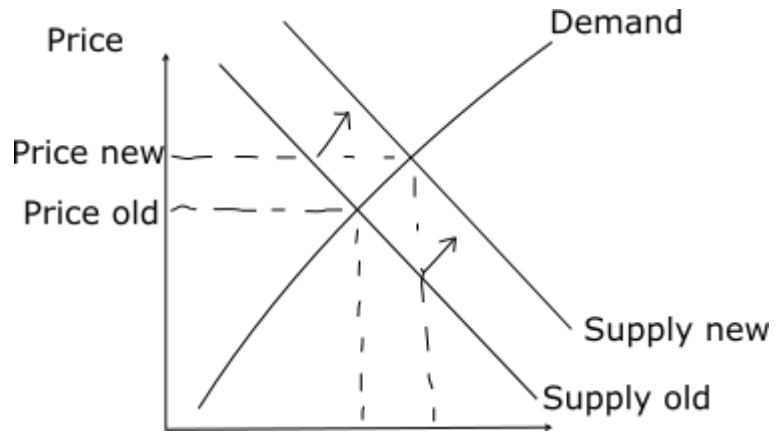
[1 point for supply surplus as consequence]

Item Info

Short: DP05

Graph: Supply and Demand

Subcomponent: Identifying mistakes



Students were asked to draw a supply and demand graph to show how a poor harvest affects the market for strawberries (strawberries being a normal good). One student drew the graph shown.

Evaluate the graph.

[For precise coding and anchor examples -> see coding manual]

[2 points for noticing that demand and supply are reversed]

[1 point for noticing the wrong slope for only supply or only demand]

[1 point for noticing the missing x-axis-label]

Item Info

Short: DP06

Graph: Supply and Demand

Subcomponent: Model evaluation

Using the supply and demand model of the labor market, it can be shown that a minimum wage above the market clearing wage leads to a decline in the quantity of labor demanded (and thus to unemployment). In reality, we rarely see this happening. Provide two short possible explanations for this contradiction.

[For precise coding and anchor examples -> see coding manual]

[1 point per possible explanation, up to a maximum of 2 points e.g.,

- requirements of the model are not met on the labor market
- influencing factors are excluded from the model
- ...

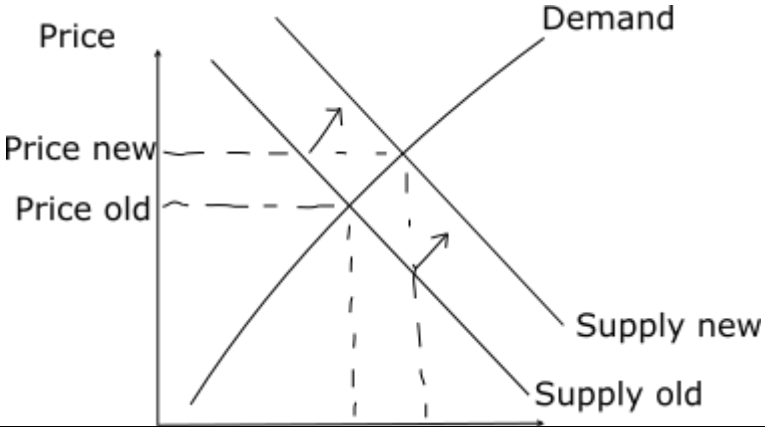
Appendix B: coding manual

Graph competence: coding manual for free response items

Supplemental material to “Measuring Graph Competence in Economic Education”

Item short	DP04
Operator	Illustration
Graph	Supply and demand
Item	Explain how the supply and demand graph can be used to outline the effect of a binding minimum price (a price above the market clearing price) on the quantity supplied and demanded.
Solution	Graphically, the minimum price corresponds to a horizontal line (parallel to the x-axis) and intersects the y-axis at the level of the government-set price. Assuming that the minimum price is binding and thus higher than the market clearing price, the straight line intersects supply and demand above the market clearing price. Accordingly, the quantity demanded at this new price is lower, but the quantity supplied is higher, resulting in a supply surplus.
Number of variables	2
Explanation of variables	DP04_01: Checks whether the test taker explains that the minimum price can be represented as a straight line. DP04_02: Checks whether the test taker explains the effect of the minimum price.

Codes for DP04_01, DP04_02			
Variable	Codes	Explanation	Examples (translated from the German Version)
DP04_01	1	The test taker explains that the minimum price can be represented as a horizontal line in the graph [Even if a wrong explanation of the effect follows afterwards]	<p>“Minimum price is entered as a horizontal straight line at the level of the corresponding price.”</p> <p>“By a horizontal line at the point on the Y axis (price).”</p> <p>“the price does not fall below a certain level (minimum price) but reaches a horizontal line there, independent of supply and demand”</p>
	0	The test taker does not explain (or incorrectly explains) that the minimum price can be represented as a horizontal line in the graph.	<p>„I do not know.“</p> <p>„Demand decreases“</p>
	-99	No answer is given	
DP04_02	1	The test taker explains the effect of a minimum price (supply higher than demand, supply gain and demand-loss, supply surplus) (“supply is higher” is not sufficient, supply and demand ratio needs to be addressed)	<p>„ A minimum price results in a supply surplus.“</p> <p>„above market price -> supply surplus“</p>
	0	The test taker does not explain the effect of minimum price sufficiently (only related to demand or only related to supply) or explains the effect incorrectly.	„ Demand decreases “
	-99	No answer is given	

Item short	DP05
Operator	Finding errors
Graph	Supply and demand
Item	<p>Students were asked to draw a supply and demand graph to show how a poor harvest affects the market for strawberries (strawberries being a normal good). One student drew the graph shown.</p> <p>Evaluate the graph.</p> 
Solution	<p>There are two different errors in the graph shown:</p> <ul style="list-style-type: none"> • The label of the x-axis (quantity) is missing. • Supply and demand are interchanged; accordingly, the supply curve should have a positive and the demand curve a negative slope. In other words, the supply should decrease (according to the sketch it increases) and the price should increase (this is correct in the sketch, but only because the supply and demand curve are swapped). <p>[The curved lines are considered as "sketch properties" not as errors].</p>
Number of variables	3
Explanation of variables	<p>DP05_01: Checks whether the test taker identifies the missing x-axis label as a mistake (and corrects it).</p> <p>DP05_02: Checks whether the test taker identifies the wrong slope of the supply curve OR identifies the increase of the supply as a mistake (and corrects it).</p> <p>DP05_03: Checks whether the test taker identifies the wrong slope of the demand curve as a mistake (and corrects it).</p>
<p>Note: Answers where it is clear that the test taker recognized that the curves are interchanged correctly argue both from a supply and demand perspective and accordingly receive a "2" in DP05_02 and DP05_03.</p>	

Codes for DP05_01, DP05_02, DP05_03			
Variable	Codes	Explanation	Examples (translated from the German Version)
DP05_01	2	The test taker identifies the missing x-axis label as a mistake and corrects it.	"Quantity is missing as the label of the x-axis".
	1	The test taker identifies the missing x-axis label as a mistake.	"x-axis is not labeled."
	0	The test taker does not identify the missing x-axis label as a mistake.	"Since supply is decreasing rather than increasing, the supply curve must shift to the left"
	-99	No answer is given	
DP05_02	2	The test taker identifies that the curves are interchanged OR that the supply curve has the wrong slope and gives the correct slope.	"Here, the supply and demand curves have been swapped." "Supply must be drawn linearly increasing and demand linearly decreasing. "
	1	The test taker identifies that the supply curve has the wrong slope without giving a correction. OR The test taker identifies that the supply shift is a mistake but gives no correction (or argues that the supply should be shifted in the other direction)	"The statement was misrepresented in the graph because a poor harvest means the supply of strawberries decreases, it does not increase as indicated in the graph."
	0	The test taker does not identify a mistake regarding supply.	"The graph clearly shows that demand remains constant, whereas the price of the supply line increases, which eventually increases the equilibrium price. This is optimal. [...]"
	-99	No answer is given	
DP05_03	2	The test taker identifies that the curves are interchanged OR that the demand curve has the wrong slope and gives the correct slope.	"You would think that demand would go down at a higher price, but in this chart, demand is greater at a higher price."
	1	The test taker identifies that the increase in demand is a mistake but gives no correction (or argues that demand should be shifted in the other direction)	[no example in the data]
	0	The test taker does not identify a mistake regarding demand.	"The statement was misrepresented in the graph because a poor harvest means the supply of strawberries is

			decreasing, not increasing as indicated in the chart."
	-99	No answer is given	

Item short	DP06
Operator	Evaluation
Graph	Supply and demand
Item	Using the supply and demand model of the labor market, it can be shown that a minimum wage above the market clearing wage leads to a decline in the quantity of labor demanded (and thus to unemployment). In reality, we rarely see this happening. Provide two short possible explanations for this contradiction.
Solution	<p>Correct approaches</p> <ul style="list-style-type: none"> - The resulting costs are covered differently/labor demand cannot be adjusted by companies at will: companies reallocate necessary costs, e.g., by increasing prices, reducing their profits, or reducing labor hours (without losing productivity) - Influencing factors are excluded from the model: wages are not the only determinant of labor demand; other factors include societal pressures to maintain jobs - The labor market differs from the goods market/assumptions of the model (e.g., perfect competition) are not given - Minimum wage affects aggregate demand: an increase in purchasing power due to minimum wage increases demand which offsets the effect <p>Wrong approaches:</p> <ul style="list-style-type: none"> - Individuals/workers are more likely to work/want to work if they are paid better, migration of workers to other sectors, work-life balance [this is right, but the reasoning refers to the household side, and in the model, the household side is not the limiting factor] - Arguments related to the fact that the minimum wage can be circumvented [not relevant]. - In certain areas, there is no qualified personnel, therefore, demand is unabatedly high [then the minimum wage would not be binding] - General model criticism: model and reality do not match [it must be explained why this is not the case]
Number of variables	1
Explanation of variables	DP06_01: Checks how many correct explanations the test taker gives. (Up to a maximum of two)

Codes for DP06_01			
Variable	Codes	Explanation	Examples (translated from the German Version)
DP06_01	2	Two correct explanations, which refer to different basic arguments (see note at the bottom)	<p>Explanation 1: Despite the increased minimum wage, workers are needed. Instead of laying off employees, a company will balance the ratio via an increased price.</p> <p>Explanation 2: Due to a higher minimum wage, the demand for goods is higher because the purchasing power is stronger. This benefits businesses."</p>
	1	One correct explanation OR one correct and one incorrect explanation OR Two explanations which refer to the same basic argument (see note at the bottom)	"When more wages are paid, more can be consumed on the part of the worker/employee, which can increase demand and bring more profit to companies."
	0	No correct explanation	"On the one hand, more unemployed people may have decided to accept a job that they had previously rejected (motivation)."
	-99	No answer is given	
<p>Note: Answers that refer to the same basic argument are generally counted as one explanation.</p> <p>Example:</p> <ol style="list-style-type: none"> 1. the minimum wage is higher than the previous wage for some jobs, but companies or employers can afford this difference without any problems in most cases. 2. companies or employers cut working hours and make part-time workers more attractive in order not to have to pay more in the end but demand higher productivity. <p>Counts as one explanation, because in both cases, companies find other sources to settle the additional cost</p>			

Appendix C: Auxiliary results

Table C1: Interrater reliability for all open-response items

Graph	Subcomponent	Short	Maximum credit	Kappa	Agreement in %	ICC
Supply & Demand	Illustration	DP04	2	0,66	85,65	0,78
Supply & Demand	Finding errors	DP05	3	0,62	75,31	0,95
Supply & Demand	Evaluation	DP06	2	0,46	68,93	0,77
Cost Function	Illustration	DK04	1	0,83	93,9	0,91
Cost Function	Finding errors	DK05	2	0,91	93,08	0,99
Cost Function	Evaluation	DK06	2	0,39	65,7	0,67
Indifference curve	Illustration	DI04	2	0,48	76,23	0,7
Indifference curve	Finding errors	DI05	3	0,77	84,96	0,82
Indifference curve	Evaluation	DI06	2	0,42	69,26	0,7

Notes: This table reports different measures of interrater reliability of two raters for all open response items used in the instrument.

Table C2: Differential Item Functioning

Itemno.	<i>Gender</i>		<i>Teacher training</i>		<i>Economics in school</i>	
	Uniform	Non-uniform	Uniform	Non-uniform	Uniform	Non-uniform
	R ² [Cat]	R ² [Cat]	R ² [Cat]	R ² [Cat]	R ² [Cat]	R ² [Cat]
1	0 [A]	0 [A]	0.001 [A]	0.001 [A]	0.001 [A]	0.001 [A]
2	0 [A]	0.01 [A]	0 [A]	0 [A]	0 [A]	0 [A]
3	0.003 [A]	0.02 [A]	0.005 [A]	0.006 [A]	0.005 [A]	0.006 [A]
4	0 [A]	0.001 [A]	0.019 [A]	0.005 [A]	-	-
5	0.001 [A]	0.002 [A]	0 [A]	0 [A]	0 [A]	0 [A]
6	0.001 [A]	0.001 [A]	0.002 [A]	0.015 [A]	0.002 [A]	0.015 [A]
7	0.01 [A]	0.003 [A]	0 [A]	0 [A]	0 [A]	0 [A]
8	0.003 [A]	0 [A]	0 [A]	0 [A]	0 [A]	0 [A]
9	0 [A]	0.012 [A]	0.001 [A]	0.005 [A]	0.001 [A]	0.005 [A]
10	0 [A]	0.016 [A]	0.007 [A]	0.006 [A]	0.007 [A]	0.006 [A]
11	0.003 [A]	0.001 [A]	0.003 [A]	0.005 [A]	0.003 [A]	0.005 [A]
12	0.012 [A]	0.002 [A]	0.001 [A]	0 [A]	0.001 [A]	0 [A]
13	0 [A]	0.002 [A]	0 [A]	0 [A]	0 [A]	0 [A]
14	0 [A]	0 [A]	0.001 [A]	0.001 [A]	0.001 [A]	0.001 [A]
15	0.004 [A]	0 [A]	0.002 [A]	0.02 [A]	0.002 [A]	0.02 [A]
16	0.012 [A]	0.001 [A]	0 [A]	0.001 [A]	0 [A]	0.001 [A]
17	0.001 [A]	0.008 [A]	0.015 [A]	0 [A]	0.015 [A]	0 [A]
18	0.012 [A]	0.001 [A]	0.001 [A]	0.001 [A]	0.001 [A]	0.001 [A]

Notes: This table reports results (Pseudo R²) for uniform and nonuniform Differential Item Functioning based on the ordinal logistic regression approach described in Zumbo (1999) and integrated into the statistical software environment R by Choi et al. (2011). Classification of DIF effects is based on the method proposed by Zumbo and Thomas (1997), where $R^2 < 0.13$ indicate negligible DIF (category A). Items with $0.13 < \Delta R^2 \leq 0.26$ are classified as moderate DIF (category B) and items with $\Delta R^2 > 0.26$ are flagged as severe DIF (category C). Focal group for Gender is female, for the variable teacher training respondents who are enrolled in a teacher training program, and for variable “Economics in school” respondents who received economic education in school. Due to missing observations, no DIF effects in item 4 are estimated for variable “Economics in school”.

Table C3: Bivariate correlations

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) Graph competence										
(2) Economic knowledge	0.30***									
(3) Economic interest	0.34***	0.28***								
(4) Age (years)	-0.01	0.05	-0.08							
(5) Female (1/0)	-0.14***	-0.18***	-0.29***	-0.05						
(6) Math grade	0.24***	0.19***	0.01	-0.19***	-0.02					
(7) German grade	0.07	0.13**	-0.12*	-0.15***	0.11**	0.35***				
(8) Econ school (1/0)	0.13**	0.03	0.26***	0.04	0.01	-0.02	-0.09*			
(9) Semester	0.14***	0.13**	-0.10	0.59***	-0.03	-0.04	-0.08*	0.03		
(10) Teacher training (1/0)	-0.07	-0.03	0.08	0.01	0.04	-0.08	-0.08	0.14***	-0.01	
(11) Econ university (1/0)	0.17***	0.10*	0.56***	-0.06	-0.07	0.01	-0.09*	0.27***	-0.05	0.27***

Notes: This table reports bivariate correlations (Pearson's r) between all variables listed in Table 1. Graph and economic competence scores are based on the IRT model described in chapter 3. Scores on economic interest are extracted from a Principal Component Analysis (PCA).