



Improving economics teaching  
and learning for over 20 years

# **Bridging the Quantitative Skills Gap: Teaching Simple Linear Regression via Simplicity and Structured Replication**

## **CONTENTS**

- 1. Introduction
- 2. Teaching SLR: Simplified examples and structured replication
- 3. A case study
- 4. Evaluation via Pedagogical Research
- 5. Related resources
- 6. Concluding remarks
- Related resources
- References
- Notes

**Steve Cook, Swansea University**

s.cook at swan.ac.uk

**Peter Dawson, University of East Anglia**

Peter.Dawson at uea.ac.uk

**Duncan Watson, University of East Anglia**

Duncan.Watson at uea.ac.uk

**Edited by Caroline Elliott, University of Warwick**

**PUBLISHED**

**February 2025**

<https://doi.org/10.53593/n4229a>

The quantitative skills (QS) gap among graduates is a pressing challenge highlighted in numerous reports emphasising the growing demand for data expertise in the workforce.

This chapter addresses this issue by focusing on the teaching of Simple Linear Regression (SLR), a prominent feature of quantitative methods training and a cornerstone of econometrics education.

To tackle the QS gap, we propose a pedagogical framework that employs simplified, bespoke examples within a replication-based approach to support the teaching and learning of SLR. By simplifying arithmetic complexity, the method emphasises the core principles of SLR, enhancing both conceptual understanding and student engagement. An evaluation based on consideration of key pedagogical issues such as cognitive load theory, active learning, self-efficacy, and technology-enhanced learning (TEL) provides strong, and often unexpected, support for the proposed approach.

Despite championing 'simplicity', the use of real-world data, as frequently prioritised in quantitative methods education, is not rejected. Instead, a balanced and context-sensitive integration of simplified and real-world datasets is suggested. This balance allows students to develop essential skills in data literacy and quantitative reasoning, creating a strong foundation for advanced learning. The proposed approach not only addresses the QS gap but also contributes to advancing effective teaching practices in economics education.

## 1. Introduction

The quantitative skills (QS) gap is a critical issue revealed through analyses across the social sciences. Studies such as the [British Academy's \*Society Counts\* \(2012\)](#) highlight systemic deficiencies in QS education, which hinder the ability of graduates to engage effectively with data. [MacInnes et al. \(2016\)](#) and [Mansell \(2015\)](#) link these gaps to reduced employability and a lack of preparedness for addressing contemporary challenges. These findings underline the pressing need for disciplines within the social sciences to improve their QS provision to better equip students for data-driven contexts. While this is reflected in discipline-specific studies for criminology, psychology and sociology (see [Chamberlain, 2016](#); [Counsell et al., 2016](#); [Scott Jones and Goldring, 2015](#)), economics faces similar challenges despite being traditionally considered a leader in QS training in the Social Sciences.

As [Mason et al. \(2015\)](#) and [Cook and Watson \(2023a\)](#) note, there is scope for development in QS education, particularly at the undergraduate level. However, economics is well-positioned to both address its own deficiencies and lead improvements across the social sciences. For instance, broader analyses such as [Quantifying the UK Data Skills Gap \(DSIT & DDCMS, 2021\)](#) and the [World Economic Forum's \*Data Science in the New Economy\* \(2019\)](#) emphasise the economic significance of data literacy and statistical reasoning in driving innovation and growth- areas in which economics is uniquely equipped to excel.<sup>[1]</sup>

A key aspect of addressing the QS gap is equipping students with a strong understanding of foundational quantitative tools. An undeniably prominent element in quantitative methods training across the social sciences, which acts as a gateway to developing critical quantitative reasoning skills, is the Simple Linear Regression (SLR) model. In economics, it plays a particularly pivotal role as both a cornerstone of econometric analysis and an essential bridge between theory and practice. By enabling the study of relationships between variables, SLR provides the basis for more advanced quantitative methods that are vital to data-driven decision-making. Its importance is further underscored by its inclusion in the [World Economic Forum's \(2019\)](#) six clusters of

data science skills, highlighting its relevance across disciplines. In this chapter, a simplified, replication-based approach to the teaching and learning of SLR is proposed, aimed at minimising unnecessary complexity while maximising accessibility and engagement.

The proposed approach to the teaching of SLR seeks to extend traditional methods and address the QS gap. The intention is to increase the focus on, and hence support the learning of, core statistical concepts by avoiding overly complex examples and employing replication-based activities. This chapter can therefore be viewed as introducing several novel contributions to enhance understanding and engagement:

1. ***The extent of simplification*** – The approach minimises arithmetic complexity by using small samples and bespoke artificial data. Statistical components such as coefficient estimates,  $R^2$ , TSS, and RSS are rounded to ensure students can focus on understanding core principles rather than grappling with cumbersome calculations.
2. ***The use of replication*** – A threefold replication framework (see [Cook and Watson, 2023a](#)) encourages students to reproduce findings using bespoke datasets. By interacting directly with the mechanics of SLR, students can deepen their understanding while actively engaging with the material.
3. ***Inclusive pedagogical evaluation*** – The chapter evaluates the simplified approach against prominent themes in pedagogical research, including cognitive load theory, active learning, self-efficacy, and technology-enhanced learning (TEL). It explores how simplicity addresses the QS gap while responding to criticisms surrounding realism and TEL.
4. ***Resource provision and spin-offs*** – The approach is designed to be flexible and scalable. It includes illustrative examples, an accompanying Excel file with worked solutions, and contributions to the Economics Network Ideas Bank — equipping educators with tools that can be adapted to diverse teaching contexts.

By placing simplified, bespoke data at the heart of SLR instruction, this chapter aims to create a focused and accessible pathway for students. This approach not only addresses foundational challenges in QS education, but also equips learners with the skills needed to progress toward more complex econometric topics and real-world applications.

While championing simplicity, this chapter acknowledges potential criticisms that it may face. A primary challenge is that simplified examples may fail to satisfy repeated calls for integrating real-world data in QS education. In econometrics, the use of real-world, often long-run data is well established ([Cook, 2016](#); [Hendry, 2015](#); [Hendry and Mizon, 2016](#); [Hendry and Nielsen, 2010](#)). Similarly, the GAISE report ([Carver et al., 2016](#)) champions the use of real-world data as one of six key recommendations for statistics education, with [Legacy et al. \(2024\)](#) highlighting its widespread adoption. Additionally, [Neumann et al. \(2013\)](#) argue that real-world data aids understanding of abstract concepts and demonstrates relevance. Beyond this, realism is central in problem-based learning (PBL), as [Hmelo-Silver \(2004, p.236\)](#) notes, '*PBL is well suited to helping students become active learners because it situates learning in real-world problems*'. Realism also features prominently in discussions of authentic assessment ([Villarroel et al., 2017](#)).

Another criticism arises from the use of pen-and-paper exercises in this approach, which may appear to conflict with the literature on TEL in Higher Education. Studies such as

[Kirkwood and Price \(2013\)](#) trace the integration of technology in education since the 1990s, while [Goodchild and Speed \(2018\)](#) describe the late 20th century as '*Epoch 2*' due to the proliferation of personal computers. [Harasim \(2000, p. 42\)](#) highlights the '*enormous innovation and expansion in online education*' during the 1980s and 1990s, and [Williams and Wong \(2009\)](#) discuss a mini-revolution in educational technology before the 21st century. Furthermore, pedagogical concepts such as Digital Natives ([Prensky, 2001a, 2001b](#)), Homo Zappiens ([Veen and Vrakking, 2006](#); [Veen and van Staaldunin, 2010](#)), and the Net Generation ([Tapscott, 1998](#)) argue for increased technology use to meet the needs of modern students.

Faced with these criticisms, this chapter does two things. First, it explicitly clarifies the proposed approach, emphasising how simplification and replication complement each other. Simplification minimises unnecessary complexity, enabling students to focus on core methods and concepts. Replication, meanwhile, provides a structured framework to reinforce learning. Second, noting previously discussed potential criticisms in relation to realism and TEL, the proposed approach is subject to a broader evaluation based on prominent issues in pedagogical research. As will be demonstrated, some criticisms lack the intuitive weight attributed to them, while other themes in pedagogical research directly support the simplified approach. Importantly, while this pedagogical evaluation indicates that potential criticisms arising from a failure to employ real-world data do not carry the weight intuitively expected, the use of real-world data is not rejected. Rather, it is argued that simplified datasets and real-world applications should coexist in teaching econometrics with their use being context-dependent. However, 'context-dependence' here is not simply in the form of simplified data being employed at an introductory level and real-world data being utilised at more advanced levels. While simplified data and examples can support the building of foundational understanding as argued here, their use can prove similarly beneficial at more advanced levels also by reducing complicating factors and allowing an increased focus on relevant issues- see, for example, [Cook et al. \(2024a\)](#) where tailored examples are employed to consider unit root testing and multivariate cointegration analysis. Similarly, while real-world data can add value and perhaps have increased relevance at more advanced stages of study, its use at more introductory levels can be beneficial to illustrate the complexity of economics data and the application of methods (see, for example, [Hendry and Nielsen, 2010](#); [Reade, 2007](#)).

The chapter is organised as follows. [Section 2](#) outlines the foundational concepts of simple linear regression and explains the replication-based approach used to teach and support understanding of this material. [Section 3](#) presents a case study demonstrating the simplified approach in practice, offering various options to highlight its flexibility and adaptability. [Section 4](#) evaluates the approach against key themes in pedagogical research, showing how simplified examples can help bridge the QS gap. [Section 5](#) discusses related resources available from the Economics Network Ideas Bank and their relationship to this chapter. Finally, [Section 6](#) offers concluding remarks, reflecting on the broader contributions of this approach to quantitative methods education.

## 2. Teaching SLR: Simplified examples and structured replication

The SLR model is a pillar of quantitative methods education, forming the foundation of econometrics provision and serving as both a critical building block and a pathway for advanced study. Its delivery typically includes well-established components: plotting a straight line imposed on a scatter graph; explaining Ordinary Least Squares (OLS)

estimation and its minimisation of the residual sum of squares; deriving OLS estimators using partial differentiation; outlining the assumptions of the linear regression model; interpreting key output statistics; and conducting hypothesis testing. While these elements are fundamental, they may often appear abstract or overly formulaic to students, particularly when concepts such as coefficient estimators, standard errors, t-statistics, and the  $R^2$  are presented with a strong algebraic and mechanical focus. These challenges in accessibility can undermine learner engagement[2] and hinder understanding of the purpose and application of SLR. A possible response to this may be to incorporate real-world data into delivery to make the material more relatable and relevant. However, this chapter proposes an alternative: using simplified examples with artificial datasets to teach the foundational aspects of SLR. By removing the complexities inherent in real-world data, this approach enables learners to focus more effectively on core statistical principles and their underlying concepts. Rather than viewing simplified examples and real-world data as mutually exclusive, we argue that their use should depend on the teaching context. Simplified examples are particularly valuable for introducing foundational concepts, as they reduce unnecessary distractions while promoting clarity and understanding. Later sections will demonstrate how these two approaches can complement each other to enhance both engagement and comprehension in the teaching of SLR.

A key feature of our approach is the use of simplified examples integrated into a replication-based framework. While debates on the importance of replication within research are well-established and have generated a voluminous literature (see [Ioannidis, 2005](#); [Clemens, 2017](#); [NASEM, 2019](#)), recent attention has focused on the benefits of incorporating replication within teaching (e.g. [Janz, 2016](#); [Stojmenovska et al., 2019](#); [Smith et al., 2021](#)). These discussions primarily emphasise the reproduction of research findings as a teaching tool. However, our approach extends this idea by adopting the broader threefold categorisation of replication proposed by [Cook and Watson \(2023a\)](#). This framework comprises three forms of replication, each serving distinct educational purposes:

1. **Direct replication**– Students are tasked with reproducing provided results exactly as they are, helping to build confidence in their ability to follow structured analyses.
2. **Step replication**– Students reconstruct results by completing hidden intermediate steps, deepening their understanding of the methods and processes underpinning the final outcomes.
3. **Flexible replication**– Students modify the provided resources to generate additional results, fostering independent thinking and adaptability in data analysis.

Together, these forms of replication create a comprehensive framework that enhances the learning experience. This progression allows students to move from foundational comprehension to developing a more adaptable skill set, reinforcing both theoretical understanding and practical abilities.

The combination of highly simplified examples and this replication framework provides a powerful approach to teaching the foundational aspects of SLR. By removing unnecessary complexities, students can focus on the core principles of linear regression while actively engaging with the material. This dual emphasis on simplicity and replication not only improves conceptual clarity, but also helps students develop practical

skills in data analysis, preparing them for more advanced topics. To demonstrate this approach, we now present an example of its application in practice.

### 3. A case study

To illustrate our approach to teaching of SLR, we present a case study alongside examples of its application. Simplicity is central to our method, both in terms of the sample size and the nature of the data used. This is exemplified by the following data set for  $\{x, y\}$ , which employs an intentionally small and unrealistic sample size ( $n = 5$ ) and consists of 'simple' values (i.e. small integers) that yield easily interpretable results for subsequent calculations:

$$x = \{3, 0, 2, 1, 4\}; \quad y = \{15, 3, 11, 7, 14\}$$

Using these data, we estimate a SLR model,  $y_i = \beta_0 + \beta_1 x_i + e_i$ , and examine the results it provides. These results, presented in Table One, include all key elements of SLR: coefficient estimates, standard errors, t-ratios, standard error of the regression, calculated information criteria, among others.<sup>[3]</sup> From inspection of the table, it is apparent that the simplified approach retains all foundational components of SLR while enhancing accessibility and clarity.

Table One				
Dependent Variable: y				
Method: Least Squares				
Sample: 15				
Included observations: 5				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.000000	1.414214	2.828427	0.0663
X	3.000000	0.577350	5.196152	0.0138
R-squared	0.900000	Mean dependent var		10.00000
Adjusted R-squared	0.866667	S.D. dependent var		5.000000
S.E. of regression	1.825742	Akaike info criterion		4.331024
Sum squared resid	10.00000	Schwarz criterion		4.174799
Log likelihood	-8.827561	Hannan-Quinn criter.		3.911732
F-statistic	27.00000			

Prob(F-statistic)	0.013847
-------------------	----------

The above data, along with the results in Table One, serve as the primary resource for our case study. The next consideration for instructors is how to use these materials effectively to teach SLR. Here, we outline four possible options, while acknowledging that lecturers may adapt these or explore alternative approaches to suit their specific teaching needs.

### Option 1

The data and Table One are provided to students, who are then tasked with reproducing the table using econometrics software. This requires students to input the data and select the appropriate commands or options to automatically generate the results. By doing so, students engage in an exercise of *direct* replication, strengthening their ability to follow structured procedures and interpret the outputs accurately.

### Option 2

The data and Table One are provided to students, who are then tasked with reproducing some or all of the results in a 'step-by-step' manner. This can be done using either a pen-and-paper approach or software such as Excel to perform the necessary calculations. Unlike results generated automatically by econometric software, this method requires students to manually compute the underlying elements. For instance, students might calculate coefficient estimates and associated standard errors, deriving intermediate values such as means and sums of products as part of the process. This constitutes an exercise in *step* replication. To illustrate this approach, Table Two and its accompanying bullet points outline how to derive the required results using Excel or a pen-and-paper method. Given the rounded and simplified nature of the data, the calculations in Table Two remain accessible, with sums of products taking straightforward values such as 10, 100, 30 and 90, and the means of the two variables being  $\bar{x} = 2$ ,  $\bar{y} = 10$ . The complete calculation of all output elements, including the p-values for test statistics, is further detailed in the accompanying Excel file, [Simple.xlsx](#).

**Table Two**

$x_i$	$y_i$	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$	$\hat{e}_i$	$\hat{e}_i^2$	$\hat{y}_i$	$(\hat{y}_i - \bar{y})^2$
3	15	1	5	1	25	5	2	4	13	9
0	3	-2	-7	4	49	14	-1	1	4	36
2	11	0	1	0	1	0	1	1	10	0
1	7	-1	-3	1	9	3	0	0	7	9
4	14	2	4	4	16	8	-2	4	16	36
<b>SUMS →</b>		0	0	10	100	30	0	10	50	90

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{30}{10} = 3$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 10 - (3 \times 2) = 4$$

$$s. e. (\hat{\beta}_0) = \sqrt{\frac{\hat{\sigma}^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}} = \sqrt{\frac{\left(\frac{10}{3}\right) \times 30}{5 \times 10}} = \sqrt{2} = 1.414 \text{ (3 d. p.)}$$

$$s. e. (\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} = \sqrt{\frac{\left(\frac{10}{3}\right)}{10}} = \sqrt{\frac{1}{3}} = 0.577 \text{ (3 d. p.)}$$

### Option 3

A redacted version of Table One is provided to students, omitting certain data. Students are then tasked with deriving the missing elements in a step-by-step manner using either a pen-and-paper approach or software, such as Excel, to perform the required calculations. This exercise can range from straightforward tasks, such as calculating t-ratios from given estimated coefficients and standard errors, to more complex problems designed to test deeper understanding. For instance, students could be assigned a challenging exercise like Question One, paired with the accompanying redacted Table Three.

### Question One

Using OLS, an investigator estimated the model  $y_i = \beta_0 + \beta_1 x_i + e_i$ , obtaining the results presented in Table Three where several calculated values have been removed and replaced with '?'. Drawing upon the results provided below, can the null hypothesis  $H_0: \beta_1 = 0$  be rejected at the 10% level of significance against the alternative  $H_1: \beta_1 \neq 0$ ?

Table Three				
Dependent Variable: y				
Method: Least Squares				
Sample: 1 5				
Included observations: 5				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.000000	1.414214	2.828427	0.0663
x	?	?	?	?
R-squared	0.900000	Mean dependent var	10.00000	
Adjusted R-squared	?	S.D. dependent var	5.000000	
S.E. of regression	1.825742	Akaike info criterion		?
Sum squared resid	?	Schwarz criterion		?



Log likelihood	-8.827561	Hannan-Quinn criter.	3.911732
F-statistic	?		
Prob(F-statistic)	?		

In this exercise, the student is tasked with assessing the significance of a coefficient- a routine task were the relevant p-value provided. However, neither the p-value nor the relevant t-ratio is available, and even the estimated coefficient and its associated standard error are omitted.[4] As a result, the 'direct' information typically required to address the question is missing. Instead, the student must rely on 'indirect' information, synthesising the available output to derive an answer. This approach presents a significant challenge, encouraging deeper and analytical thinking and problem-solving skills. Here is one approach to addressing the question:

- The question involves consideration of the single slope coefficient ( $\beta_1$ ) in the estimated model. As the  $R^2$  is present, the F-test for  $R^2 = 0$  can be used to provide a test of the significance of the single slope coefficient. The square root of this F-statistic provides the absolute value of the t-ratio we require. We can denote this t-ratio as  $t_{\beta_1}$ .
- The value of  $|t_{\beta_1}|$  can be compared to the reported t-ratio ( $t_{\beta_0}$ ) for  $H_0: \beta_0 = 0$  against  $H_1: \beta_0 \neq 0$ . Comparison of the absolute value of  $t_{\beta_1}$  with the reported  $t_{\beta_0}$  will then determine whether the relevant p-value we require is less (or greater) than the reported 6.63% associated with  $t_{\beta_0}$ . We can do this as the same degrees of freedom will be used for distribution for the two t-ratios.
- Drawing these two steps together, we can arrive at an answer as follows:

$$- F_{(R^2=0)} = \frac{R^2}{1-R^2} \times \frac{n-k-1}{k} = \frac{0.9}{0.1} \times \frac{3}{1} = 27$$

- Note we have again chosen numbers to make the calculation straightforward (i.e. 3, 9, 27).

$$- t_{\beta_1} = \pm \sqrt{27} = \pm 5.196 \text{ (3 d.p.)}; \therefore |t_{\beta_1}| = 5.196$$

- As  $|t_{\beta_1}| > |t_{\beta_0}|$ , we can conclude that the p-value associated with  $t_{\beta_1}$  is less than the p-value reported for  $H_0: \beta_0 = 0$  against  $H_1: \beta_0 \neq 0$ . As the latter p-value is 6.63%, the p-value for  $H_0: \beta_1 = 0$  against  $H_1: \beta_1 \neq 0$  is smaller than 6.63%. Therefore  $H_0: \beta_1 = 0$  against  $H_1: \beta_1 \neq 0$  can be rejected at the 10% level of significance.

The arithmetic is straightforward: the required '27' is easily calculated, its square root clearly exceeds 5, and this value is evidently greater than the t-statistic of 2.828, to which it must be compared. However, while the numerical calculations are simplified, the real challenge lies in recognising and appreciating the logical steps needed to arrive at the solution.

After completing this task, learners can either be shown the output from Table One or asked to generate this themselves once the data are subsequently provided, revealing that the actual p-value is 1.38%. This value can then be discussed to highlight that it not only supports rejection of the null hypothesis at the 10% significance level but also at the more stringent 5% level. Naturally, this question can be further adapted. For example,

Question Two extends the challenge by requiring learners to engage with concepts such as coefficient signs, one-tailed testing, and manipulation of p-values.

### Question Two

Using OLS, an investigator estimated the model  $y_i = \beta_0 + \beta_1 x_i + e_i$  obtaining the results presented in Table Three where several calculated values have been removed and replaced with '?'. The investigator also finds that  $\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = 30$ . Drawing upon these results, can the null hypothesis  $H_0: \beta_1 = 0$  be rejected at the 5% level of significance against the alternative  $H_1: \beta_1 > 0$ ?

Table Three				
Dependent Variable: y				
Method: Least Squares				
Sample: 1 5				
Included observations: 5				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.000000	1.414214	2.828427	0.0663
x	?	?	?	?
R-squared	0.900000	Mean dependent var	10.00000	
Adjusted R-squared	?	S.D. dependent var	5.000000	
S.E. of regression	1.825742	Akaike info criterion	?	
Sum squared resid	?	Schwarz criterion	?	
Log likelihood	-8.827561	Hannan-Quinn criter.	3.911732	
F-statistic	?			
Prob(F-statistic)	?			

Once again, no 'direct' information is available for the coefficient under consideration: there is no estimated value, standard error, t-ratio, or p-value. However, a solution can be derived by applying the approach used in Question One and incorporating the knowledge that  $\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = 30$ . Since this cross-product term is positive,  $\hat{\beta}_1$  must be positive. Thus, we can specify  $t_{\beta_1} = 5.196$ , rather than  $t_{\beta_1} = \pm 5.196$  as noted in Question 1 earlier, following recognition of the sign of  $\hat{\beta}_1$ . With the two-tailed p-value for  $H_0: \beta_1 = 0$  below 6.63%, the one-tailed p-value will be less than 3.315% (= 6.63% ÷ 2, to 3 d.p.). Consequently, the one-tailed test results in rejection at the 5% significance level. To reinforce these findings, students can either be provided with the output or tasked with

generating it automatically using econometrics software. Alternatively, they can calculate the p-value manually using tools such as Excel, enabling further discussion.

#### Option 4

The data and Table One are provided to students, who are then tasked with manipulating the series and explaining the resulting outputs. A straightforward example involves scaling the data.<sup>[5]</sup> For instance, students could be asked to double the values of the dependent variable and/or regressor and explain the relationship between original (Table One) and the revised outputs. Econometric software can be used to automate this process or the required calculations can subsequently be reproduced manually using Excel or a pen-and-paper approach. This constitutes an exercise in *flexible* replication, as students manipulate the provided resources to generate additional results.

Having demonstrated the use of simplification and structured replication in teaching SLR, along with the flexibility it provides, we now evaluate this approach in relation to prominent themes in pedagogical research.

## 4. Evaluation via Pedagogical Research

This section provides a pedagogical evaluation of the simplified approach advocated in this chapter by examining its alignment with key debates and discussions in the pedagogical literature. It will be argued that, for certain educational challenges, the simplified approach receives strong, and perhaps self-evident, support; for instance, cognitive load theory endorses it due to its elimination of extraneous cognitive load. In other cases, the benefits of the approach exceed initial expectations. Despite its simplicity, it facilitates active learning by generating exercises that present significant cognitive challenges through the requirement for complex synthesis. However, when addressing issues such as real-world data and TEL, we demonstrate that some pedagogical concerns require careful interpretation.

### 4.1. Active Learning

Earlier discussion highlighted the issue of engagement, which naturally leads to the consideration of active learning and its transformative impact on educational practice. Active learning represents a shift away from traditional lecture-based teaching, focusing instead on directly involving students in the learning process through interactive methods (MacManaway, 1970; Bonwell and Eison, 1991; Prince, 2004). These methods include activities such as discussions, problem-solving, and case studies, all designed to enhance engagement and deepen students' understanding of the material. However, it is overly simplistic to define active learning as merely incorporating 'some activity' into the classroom. Mayer (2021), as discussed by Cook and Watson (2023b, c), categorises active learning within a '2x2' framework that distinguishes between levels of behavioural and cognitive activities. The framework identifies four potential outcomes:

- {low cognitive, low behavioural} = ineffective, passive instruction
- {low cognitive, high behavioural} = ineffective, active instruction
- {high cognitive, low behavioural} = effective, passive instruction
- {high cognitive, high behavioural} = effective, active instruction

Therefore, while sufficient behavioural activity supports active learning, its effectiveness ultimately depends on the level of cognitive activity involved. The use of simplification and structured replication in teaching SLR provides a robust framework for integrating active learning effectively. By presenting examples, exercises can be designed to vary in their 'behavioural' and 'cognitive' demands. For instance, classroom discussions can incorporate tabulated examples where coefficient estimates and standard errors illustrate the construction of t-ratios, or sums of products and means demonstrate the derivation of coefficient estimates. Alongside these more 'substitution-based' exercises, which primarily promote behavioural engagement, more challenging tasks like those discussed in previous questions require higher levels of cognitive activity. These tasks encourage deeper understanding by requiring synthesis and application of information.

The simplified approach here can therefore support active learning in various forms. The benefits of this are further amplified by the advantages of incorporating active learning in large groups (see [Jerez et al., 2021](#)), particularly relevant as SLR is often taught in large, compulsory introductory econometrics modules.

## 4.2. Problems and The Expertise Reversal Effect

Our discussion of simplified examples naturally connects to an extensive literature on examples, problem-solving and instruction. Within this body of research, two prominent themes emerge, differing on whether instruction should *precede* or *follow* problem-solving. Research advocating for problem solving before instruction often focuses on the concept of 'productive failure' (see, inter alia, [Kapur 2008, 2012, 2015](#)). Productive failure introduces an initial 'invention phase', where learners tackle novel problems before receiving instruction. This approach aims to activate prior knowledge and enhance the effectiveness of subsequent teaching by leveraging failure as a learning opportunity.<sup>[6]</sup> Additional research in this area explores various dimensions, such as enhancing problem-solving before instruction through contrasting cases and metacognitive (non-topic-specific) prompts ([Loibl et al., 2017](#); [Roll et al., 2012](#)), extending productive failure beyond STEM disciplines ([Nachtigall et al., 2020](#)), and investigating the role of collaboration in its effectiveness ([Brand et al., 2023](#)). Conversely, other research argues against minimal guidance during early learning stages ([Kirschner et al., 2006](#)), advocating for worked examples over pure problem-solving (see, inter alia, [Kalyuga et al., 2001, 2003](#); [Sweller and Cooper 1985](#); [Cooper and Sweller 1987](#)). This debate introduces the Expertise Reversal Effect (ERE) ([Kalyuga et al., 2003](#)), which highlights how the effectiveness of instructional methods depends significantly on a learner's prior knowledge and experience. As learners gain expertise, the need for direct instruction *diminishes*, while the benefits of engaging in more complex problem-solving tasks *increase*. This transition is reflected in research by [Kalyuga et al. \(2003\)](#), the progression from 'worked-out examples' to 'partially worked-out examples' noted by [Hadfield \(2021\)](#), and the 'fading procedure' described by [Renkl and Atkinson \(2003\)](#). Gradually increasing task complexity through these approaches fosters deeper learning while maintaining learner engagement and avoiding overwhelm. The ERE also informs discussions about guided versus discovery-based learning (see, inter alia, [Lefrancois, 1997](#); [Phillips, 1998](#); [Mayer, 2004](#)), highlighting a continuum where instructional guidance decreases as learner competence grows. This concept supports a balanced approach to teaching, where instructional methods adapt to the developing expertise of learners.

Support for the ERE can be found in Cognitive Load Theory (CLT). Rooted in the work of Sweller et al. ([1998, 2019](#)) and further developed by [van Merriënboer and Sweller \(2005\)](#), CLT distinguishes between intrinsic, extraneous, and germane cognitive loads. It advocates reducing unnecessary cognitive burden (extraneous load), while promoting

activities that enhance meaningful learning (germane load). References to CLT are evident within the ERE literature. For example, [Kirschner et al. \(2006\)](#) emphasise the distinction between working and long-term memory, describing learning as ‘*a change in long-term memory*’ (Kirschner et al., 2006, p.75) and [Renkl and Atkinson \(2003\)](#) incorporate the concepts of intrinsic, extrinsic, and germane loads in their proposal of the fading procedure.

Applying CLT to quantitative methods education requires careful consideration of how statistical concepts are presented to students. The theory suggests that simplifying the delivery of complex information can significantly enhance comprehension. For instance, using worked examples followed by gradually introducing problem-solving tasks helps to manage cognitive load. By initially reducing task complexity and focusing on core concepts, educators can prevent students from feeling overwhelmed, fostering a deeper understanding of the material. This approach aligns with the signalling theory proposed by [Mautone and Mayer \(2001\)](#), which emphasises directing students’ attention to essential information while minimising extraneous distractions. The simplified approach discussed here is supported by both CLT and the ERE. By eliminating extraneous distractions and avoiding complicated arithmetic often associated with real-world data sets, students can focus on the calculation, use, and interpretation of key elements (e.g. coefficient estimates, t-ratios, RSS,  $R^2$  etc.). As students’ understanding grows, the complexity of problems can be progressively increased, as illustrated by the two questions presented above.

### 4.3. ‘Anxiety towards quants’ and self-efficacy

In the context of quantitative methods education, research on ‘anxiety towards quants’ features prominently in the pedagogical literature. This line of inquiry dates back to early work on number anxiety by [Dreger and Aiken \(1957\)](#), with numerous subsequent studies expanding on this foundation (see, for example, [Cook, 2022](#); [Cook et al., 2019](#); [Cook and Watson, 2023a](#); [Dowker et al., 2016](#); [Huang and Mayer, 2016](#); [Huang et al., 2023](#); [Onwuegbuzie, 2004](#)). The simplified approach advocated here offers a practical means to address this anxiety by reducing the complexity of the examples used. For instance, lengthy samples of real-world data, which may exhibit challenging issues such as non-stationarity, outliers, changes in variance and breaks in trend, and which often include values recorded to many significant figures or decimal places, are deliberately avoided. Instead, we propose an approach characterised by short samples with ‘simple’ values, coefficient estimates that are single digit integers, and sums of squares such as 30 and 100. By adopting this simplified framework and avoiding additional complicating factors, potential anxiety can be mitigated, creating a more accessible learning environment.

In addressing ‘anxiety’, our approach also acknowledges the critical role of self-efficacy. Within the context of quantitative education, self-efficacy is pivotal in shaping students’ perceptions of their ability to effectively apply and interpret quantitative methods (see, for example, [Bandura, 1977](#); [Zahaciva et al., 2005](#); [Huang et al., 2020](#)). Strengthening self-efficacy can motivate students to approach quantitative problems with greater confidence, thereby enhancing their learning experience and fostering mastery of complex quantitative concepts. Intuitively, a negative relationship exists between anxiety and self-efficacy, a finding supported by research (see, inter alia, [Rozgonjuk et al., 2020](#)). As argued by [Cook and Watson \(2023a\)](#), the replication-based approach adopted here offers a valuable means of promoting self-efficacy. By setting clear targets or objectives for students to achieve, this approach enables learners to build confidence through the successful reproduction of required results. Mastery of quantitative methods is

demonstrated by achieving these objectives, fostering recognition of one's ability and reinforcing self-efficacy.

To strengthen self-efficacy in quantitative methods, instructors can design replication exercises that gradually increase in complexity as learners' skills grow. Initially, starting with simple, well-structured tasks allows helps students build a solid foundation of confidence. As they demonstrate competence with these basics, the exercises can become more challenging. This approach echoes the principles of CLT and the ERE, both of which advise easing learners into complexity to prevent overwhelm and support more meaningful learning (van Merriënboer and Sweller, 2005; Kirschner et al., 2006). Working first with simplified datasets ensures that students fully grasp fundamental statistical concepts before encountering the unpredictability of real-world data. Over time, steadily raising the level of difficulty reinforces both their understanding and their belief in their own abilities. Cook and Watson's (2023a) categorisation of replication offers a blueprint for this gradual progression. Direct replication tasks encourage students to verify given results, helping them gain immediate confidence. Moving to step replication requires them to uncover missing computations on their own, fostering deeper analytical thinking and independent problem-solving. Finally, flexible replication challenges learners to modify datasets and adapt their methods accordingly, strengthening their readiness to deal with new, unfamiliar scenarios. Together, these stages not only refine technical skills but also nurture a stronger sense of self-efficacy. By carefully sequencing the complexity of tasks, instructors prepare students to approach real-world data with greater assurance and competence.

#### 4.4. Data Literacy

Addressing the QS gap and improving data literacy remain central goals as students prepare for the demands of data-rich professional environments. The simplified approach offered here invites learners to engage directly with core statistical concepts. By performing calculations manually and then confirming their results through replication and selective software use, students gain a deeper grasp of the logic behind quantitative methods than they would if they depended solely on automated output. However, it is important to acknowledge what is not happening at this stage. Without real-world data, students miss the opportunity to learn crucial skills such as retrieving datasets from reputable sources, cleaning messy information, and, critically, linking empirical results to meaningful events or policy changes in the real economy or society.<sup>[7]</sup>

This limitation must be understood in light of the primary goal at the introductory level: establishing competence and confidence before introducing complexity. Focusing initially on simplified, artificial datasets keeps the spotlight on fundamental principles — how to compute parameters, interpret test statistics, and evaluate model fit — without the distractions of complicated data structures or uncertain variable definitions. Students gain a strong conceptual foothold. They come to understand not just how econometrics works, but why.

As learners advance, this foundational competence sets the stage for more practical applications. Once students are comfortable with the basics, instructors can gradually introduce real-world data, prompting learners to connect statistical findings with actual economic or social phenomena. For instance, they may learn how to collect and clean data on unemployment rates before and after a significant policy reform, or track how consumption patterns shift in response to changes in interest rates or government spending. Students can compare these empirical findings with theory, contemporary

debates, or historical contexts, thus bridging the gap between abstract econometric concepts and the tangible world they will eventually navigate as professionals.

In the upcoming section, we will explore strategies to incorporate genuine datasets and link quantitative outcomes to the events and policies that shape them. By carefully timing the introduction of real-world complexity, educators ensure that students are ready to appreciate the complex interplay between numbers, theories, and lived experiences. This staged approach balances the need for conceptual clarity early on with the eventual ability to interpret real data in meaningful, context-rich ways—ultimately leading to a more comprehensive and practically relevant mastery of quantitative skills.

#### 4.5. Authentic delivery

While real-world data often appears central to econometrics education, simply using them does not guarantee a truly realistic or instructive experience—particularly in the introductory teaching of SLR. Consider the challenge of modelling a consumption function using only one explanatory variable (income) with time series data. Such data are frequently non-stationary and prone to outliers, with analysis potentially complicated by issues such as omitted variables, non-linearities, and structural shifts. The single-regressor SLR model can therefore be severely compromised as a means of providing a genuine representation of actual economic relationships. Similar difficulties are not restricted to the analysis of time series data. For example, analysis of the well-known Boston housing market data of [Harrison and Rubinfeld \(1978\)](#) by [Castle et al. \(2023\)](#) required consideration of ‘506 impulse indicators, 504 step indicators, and 21 free regressors’ ([Castle et al., 2023, p.41](#)), along with regional dummies and interaction terms, to arrive at a satisfactory empirical model. This degree of intricacy illustrates that achieving ‘realism’ through application of the SLR is elusive, if not impossible.

Beyond these specific examples, the broader literature underscores these challenges. As [Castle et al. \(2023, p. 31\)](#) explain, “*Complete and correct a priori specifications almost never exist for models of observational data, so model discovery is unavoidable*”. Similarly, [Hendry \(2018, p.119\)](#) reminds us that “*all empirical macro-econometric models are non-constant, and mis-specified in numerous ways*”. Although these observations stem from advanced econometric research, they highlight why realism is hard to achieve with a basic SLR model. At the beginner’s stage, using real-world data often emphasises what SLR cannot accomplish, rather than helping students understand its foundational principles.

Instead of trying to replicate the complex reality that experienced econometricians face, it may be more effective to pursue what [Savery and Duffy \(1995\)](#) call ‘authenticity’. In this approach, authenticity does not mean mirroring professional research conditions outright. Rather, it involves selecting meaningful tasks that promote genuine understanding of fundamental concepts—estimating parameters, interpreting test statistics, and examining empirical output—without overwhelming learners with messy data. Simplified datasets enable students to engage deeply with these basics, building a solid conceptual platform before tackling advanced challenges.

Of course, none of this rules out the use of real-world data. As learners progress to more sophisticated topics - such as unit root testing, cointegration, or vector autoregressions - complex real-world datasets become invaluable. At the same time, there is scope for considering the use of real-world data at more introductory levels of analysis once foundational knowledge has been established. For example, the relevance of multiple

linear regression in the modelling of real-world data can be illustrated through its use in exploring real-world issues- for example, [Reade's \(2007\)](#) examination of the determinants of football match attendances. Similarly, while real-world data can be employed in teaching at both introductory and higher levels, the benefits of simplified artificial examples are not restricted to introductory levels alone but can also be valuable at higher levels. For example, as illustrated by [Cook et al. \(2024b\)](#), simplified artificial examples can be employed to focus attention and generate bespoke examples to support the delivery of more advanced topics. However, at the introductory stage, simplification remains key to nurturing core skills and confidence. Once students have mastered the essentials in a structured, accessible environment, they will be far better prepared to navigate the intricate landscape of advanced econometric modelling and the complexities of real-world data.

#### 4.6. Technology-Enhanced Learning

Technology-Enhanced Learning (TEL) holds a well-established place in higher education, as noted by [Kirkwood and Price \(2013\)](#), [Goodchild and Speed \(2018\)](#), [Harasim \(2000\)](#), and [Williams and Wong \(2009\)](#). At the same time, today's students, often described as Digital Natives, Homo Zappiens, or the Net Generation ([Prensky, 2001a, 2001b](#); [Veen and Vrakking, 2006](#); [Veen and van Staaldouin, 2010](#); [Tapscott, 1998](#)), are frequently portrayed as naturally demanding more technologically infused learning experiences. Given this backdrop, it might seem counterintuitive to advocate for an approach that involves substantial pen-and-paper calculation. Yet a closer examination reveals that this is less a rejection of technology than a strategic deployment of it at the right time and for the right tasks.

Criticism of a simplified, low-tech start can be viewed from both demand and supply angles. On the demand side, expectations for high-level technology use are often driven by the belief that modern students need and want digital solutions for all aspects of their learning. However, the literature shows that these assumptions deserve scrutiny. [Saunders and Gale \(2012\)](#) caution that technology is not a cure-all for educational problems. Moreover, the concept of a uniform "Net Generation" with uniform preferences and skills has been questioned by [Helsper and Eynon \(2010\)](#), [Jones et al. \(2010\)](#), and [Kirschner and van Merriënboer \(2013\)](#). Students are not a monolithic block with identical technological fluency or desires. In reality, effective pedagogy depends on using technology where it genuinely improves learning rather than simply meeting perceived external expectations.

On the supply side, the approach described here does not exclude technology. Quite the opposite: it introduces technology gradually and thoughtfully. For instance, students can first engage with basic calculations by hand, which helps them internalize core concepts such as parameter estimation, hypothesis testing, and model interpretation. Once these principles are understood, they can turn to econometric software to verify and extend their manual results. This sequencing ensures that technology is seen not as a shortcut or a black box but as a powerful tool to deepen understanding. Students learn how software complements their analytical reasoning, rather than replacing it. They gain experience in moving back and forth between the use of alternative software, manual methods and digital outputs, seeing how theoretical insights map onto computed results.

This balanced integration of TEL offers multiple benefits. It provides learners with a tangible grasp of statistical principles and a clear sense of what the technology is doing behind the scenes. By avoiding early over-reliance on automated routines, students



develop stronger problem-solving skills and a more intuitive feel for the data. When they do advance to more complex tasks- such as working with real-world datasets or exploring advanced econometric techniques- they are better prepared to use technology thoughtfully, understanding both its capabilities and limitations.

In short, our approach does not ignore the potential of TEL; it refines it. By combining hands-on learning with informed software use, it delivers a richer, more nuanced learning experience that aligns with established pedagogical goals. It respects the central role of technology in modern education, yet leverages it in a way that reinforces rather than dilutes conceptual understanding. In doing so, it meets the broader aims of active learning, cognitive load management, self-efficacy building, and anxiety reduction—showcasing how TEL, used judiciously, is not a concession to modern demands, but a meaningful enhancement to rigorous quantitative skills education.

## 5. Related resources

Numerous resources are closely related to this chapter. Motivated directly by the discussion of the ERE, Cook et al. (2025a, b) explore the development of alternative exercises to support teaching introductory econometrics. While Cook et al. (2025a) focuses on hypothesis testing and model selection, Cook et al. (2025b) adopts a similar approach to the design of exercises for diagnostic testing. In both cases, these exercises aim to tease out key concepts in introductory econometrics by adopting alternative perspectives. As noted in these studies, the more puzzling or cryptic nature of the exercises compared to conventional formats suggests their application at a later stage of the learning process. An alternative approach to designing econometrics exercises is also presented in Cook et al. (2024a) with the use of redaction advocated. Here, the intention is not to create simplistic exercises where missing elements are calculated through direct substitution of provided values, but rather to use redaction to support multi-step exercises requiring the synthesis of information and deeper understanding. Although Cook et al. (2024a) focuses on higher-level topics in econometrics, it shares the same underlying motivation as Cook et al. (2025a, 2025b) in designing exercises that diverge from traditional formats. A common feature across all three studies is the use of artificially generated data to create exercises that allow specific elements to be emphasised through bespoke results. This deliberate use of artificial data and tailored exercises aligns closely with the prominent theme championed in this chapter, demonstrating the effectiveness of such methods in enhancing learning outcomes.

The use of artificial data to create bespoke exercises is also evident in the studies by Cook and Watson (2024) and Cook et al. (2024b), where crossnumber puzzles are employed as an innovative approach to econometrics exercises. These puzzles encourage active engagement by requiring students to complete specific tasks that involve generating, evaluating, and interpreting empirical results. This approach builds on earlier work by Cook and Watson (2023c), which introduced crosswords as a learning tool for econometrics. Additional studies predating the current work, such as Cook (2022) and Cook and Watson (2023b), have similarly employed tailored exercises to provide targeted illustrations of econometric concepts.

## 6. Concluding remarks

This chapter outlines a practical and adaptable strategy for teaching Simple Linear Regression (SLR) that addresses the persistent quantitative skills (QS) gap observed in

economics and the broader social sciences. By starting with small, custom-made datasets and relying on a replication-based framework, instructors can create a focused learning environment that encourages students to understand the “why” behind statistical methods rather than becoming mired in unwieldy calculations from the start. This intentional simplification helps learners build a firm conceptual foundation, ensuring that their early interactions with econometrics foster clarity and reduce the risk of frustration or disengagement.

At the introductory stage, this approach allows educators to structure their lessons more thoughtfully. Instead of presenting large, messy datasets that can overwhelm beginners, instructors can highlight fundamental principles through straightforward examples. For instance, by changing just a few values within an artificial dataset, they can demonstrate how parameter estimates, test statistics, and p-values shift in response. Such incremental complexity ensures that students gradually internalise statistical concepts, improving their self-efficacy and lowering any anxiety they may feel toward quantitative methods. As a result, students not only learn the mechanics of SLR but also grasp the deeper logic of why certain steps are taken— a crucial step in bridging the QS gap.

This approach also lends itself to a rich variety of teaching methods and evaluative tasks. Instructors might start by having students replicate simple, fully worked-out examples to confirm that they can follow procedures accurately. As learners gain confidence, educators can introduce step-replication exercises, where key intermediate values are hidden, prompting students to work through the calculations themselves. In time, students can engage in flexible replication tasks, modifying datasets, altering variables, or exploring different parameterisations to see how these changes impact their results. Such exercises push beyond rote learning, fostering critical thinking, problem-solving, and adaptability- skills invaluable to both economists and social scientists who may initially feel less comfortable handling quantitative data.

As learners progress, educators can seamlessly transition to real-world datasets and more advanced econometric tools. By this stage, students have already mastered the core concepts of SLR and understand the reasoning behind various calculations, making them better equipped to appreciate the nuances and complexities of actual economic data. They will be ready to confront issues such as non-stationarity, outliers, omitted variables, or dynamic specifications without feeling out of their depth. The progression from simplified artificial examples to more authentic scenarios respects the learner’s cognitive development and cements the foundational knowledge needed to tackle these advanced challenges.

Ultimately, this gradual and carefully sequenced approach helps close the QS gap from multiple angles. Economics students can more easily connect theoretical material learned in lectures to empirical evidence, while learners from other social science backgrounds gain a welcoming entry point into quantitative analysis. The steady blend of pen-and-paper work, well-timed technology use, and eventual introduction of complex, real-world data ensures that students do not merely memorise techniques- they learn to reason quantitatively. In doing so, instructors prepare a generation of analysts, researchers, and policymakers who can navigate the increasingly data-driven landscapes of their respective fields with confidence, curiosity, and competence.

## Related resources

Cook, S. 2022. Cointegration and spurious regression: Enhancing delivery via replication, empirical application and simulation. *Economics Network Ideas Bank*. <https://doi.org/10.53593/n3536a>

Cook, S. and Watson, D. 2023b. A pedagogically-driven approach to teaching higher-powered unit root testing. *Economics Network Ideas Bank*. <https://doi.org/10.53593/n3591a>

Cook, S. and Watson, D. 2023c. Crosswords and the 'Active Learning' Quest. *Economics Network Ideas Bank*. <https://doi.org/10.53593/n3585a>

Cook, S. and Watson, D. 2024. Developing econometric and data analysis skills: Championing the crossnumber puzzle. *Economics Network Ideas Bank*. <https://doi.org/10.53593/n4142a>

Cook, S., Dawson, P. and Watson, D. 2024a. Teaching Econometrics: The role of the 3Rs. *Economics Network Ideas Bank*. <https://doi.org/10.53593/n4182a>

Cook, S., Dawson, P. and Watson, D. 2024b. Building multiple linear regression skills via 'puzzling' active learning. *Economics Network Ideas Bank*. <https://doi.org/10.53593/n4163a>

Cook et al. 2025a. Beyond exercises in substitution and direct interpretation: The use of 'puzzles' in the teaching of linear regression. *Economics Network Ideas Bank*. <https://doi.org/10.53593/n4223a>

Cook et al. 2025b. Alternative exercises to develop understanding of diagnostic testing. *Economics Network Ideas Bank*. <https://doi.org/10.53593/n4225a>

## References

Bandura, A. 1977. Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review* 84, 191-215. [https://doi.org/10.1016/0146-6402\(78\)90002-4](https://doi.org/10.1016/0146-6402(78)90002-4)

Bonwell, C. and Eison, J. 1991. Active learning: Creating excitement in the classroom. Washington DC: George Washington University. **ERIC Number: ED336049**

Brand, C., Hartmann, C., Loibl, K. and Rummel, N. 2023. Do students learn more from failing alone or in groups? Insights into the effects of collaborative versus individual problem solving in productive failure. *Instructional Science* 51, 953-976. <https://doi.org/10.1007/s11251-023-09619-7>

British Academy. 2012. **Society Counts: Quantitative Skills in the Social Sciences and Humanities**. London: The British Academy.

Buchele, S. 2020. Evaluating the link between attendance and performance in higher education: the role of classroom engagement dimensions. *Assessment & Evaluation in Higher Education*, 46, 132-150. <https://doi.org/10.1080/02602938.2020.1754330>

Carver, R., Everson, M., Gabrosek, J., Horton, N., Lock, R., Mocko, M., Rossman, A., Roswell, G. H., Velleman, P., Witmer, J., and Wood, B. 2016. GAISE college report ASA revision committee, "Guidelines for Assessment and Instruction in Statistics Education College Report 2016," [https://www.amstat.org/docs/default-source/amstat-documents/gaisecollege\\_full.pdf](https://www.amstat.org/docs/default-source/amstat-documents/gaisecollege_full.pdf).

Castle, J., Doornik, J. and Hendry, D. 2023. Robust discovery of regression models. *Econometrics and Statistics* 26, 31-51. <https://doi.org/10.1016/j.ecosta.2021.05.004>

Chamberlain, J. 2016. Ensuring the criminological skills of the next generation: a case study on the importance of enhanced quantitative method teaching provision. *Journal of Further and Higher Education* 41, 448-459. <https://doi.org/10.1080/0309877X.2015.1117602>

Clemens, M. 2017. The meaning of failed replications: A review and proposal. *Journal of Economic Surveys* 31, 326-342. <https://doi.org/10.1111/joes.12139>

Coners, A., Matthies, B., Vollenberg, C., and Koch, J. 2024. Data skills for everyone! (?) - An approach to assessing the integration of data literacy and data science competencies in Higher Education. *Journal of Statistics and Data Science Education*, 1-37. <https://doi.org/10.1080/26939169.2024.2334408>

Cook, S. 2016. Modern econometrics: Structuring delivery and assessment. *Cogent Economics and Finance* 4. <https://doi.org/10.1080/23322039.2016.1152705>

Cook, S. 2022. Cointegration and spurious regression: Enhancing delivery via replication, empirical application and simulation. *Economics Network Ideas Bank*. <https://doi.org/10.53593/n3536a>

Cook, S. and Watson, D. 2023a. The use of online materials to support the development of quantitative skills. In *The Handbook of Teaching and Learning Social Research Methods*, Nind, M. (ed.). Cheltenham: Edward Elgar. <https://doi.org/10.4337/9781800884274.00028>

Cook, S. and Watson, D. 2023b. A pedagogically-driven approach to teaching higher-powered unit root testing. *Economics Network Ideas Bank*. <https://doi.org/10.53593/n3591a>

Cook, S. and Watson, D. 2023c. Crosswords and the 'active learning' quest. *Economics Network Ideas Bank*. <https://doi.org/10.53593/n3579a>

Cook, S., Watson, D. and Vougas, D. 2019. Solving the quantitative skills gap: a flexible learning call to arms! *Higher Education Pedagogies* 4, 17-31. <https://doi.org/10.1080/23752696.2018.1564880>

Cooper, G. and Sweller, J. 1987. The effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology*, 79, 347-362. <https://doi.org/10.1037/0022-0663.79.4.347>

Counsell, A., Cribbie, R. and Harlow, L. 2016. Increasing literacy in quantitative methods: The key to the future of Canadian psychology. *Canadian Psychology* 57, 193-201. <https://doi.org/10.1037/cap0000056>

Cuddington, K., Abbott, K., Adler, F., Aydeniz, M., Dale, R., Gross, L., Hastings, A., Hobson, E., Karatayev, V., Killion, A., Madamanchi, A., Marraffini, M., McCombs, A., Samyono, W., Shiu, S., Watanabe, K., and White, E. 2023. Challenges and opportunities to build quantitative self-confidence in biologists. *BioScience* 73, 364-375. <https://doi.org/10.1093/biosci/biad015>

Department for Science, Innovation and Technology and Department for Digital, Culture, Media and Sport [DSIT & DDCMS]. 2021. **Quantifying the UK Data Skills Gap**. London: HMSO.

Dowker, A., Sarkar, A. and Looi, C. 2016. Mathematics anxiety: what have we learned in 60 years? *Frontiers in Psychology* 7, 508. <https://doi.org/10.3389/fpsyg.2016.00508>

Dreger R. and Aiken L. 1957. The identification of number anxiety in a college population. *Journal of Educational Psychology* 48, 344-351. <https://doi.org/10.1037/h0045894>

Ghodoosi, B., West, T., Li, Q., Torrisi-Steele, G., and Dey, S. 2023. A systematic literature review of data literacy education. *Journal of Business and Finance Librarianship* 28, 112-127. <https://doi.org/10.1080/08963568.2023.2171552>

Goodchild, T., and Speed, E. 2018. Technology enhanced learning as transformative innovation: a note on the enduring myth of TEL. *Teaching in Higher Education* 24, 948-963. <https://doi.org/10.1080/13562517.2018.1518900>

Hadfield, K. 2021 Providing ability to probability: Reducing cognitive load through worked-out examples. *Teaching Statistics* 43, 28-35. <https://doi.org/10.1111/test.12244>

Harasim, L. 2000. Shift happens: Online education as a new paradigm in learning. *Internet and Higher Education* 3, 41-61. [https://doi.org/10.1016/S1096-7516\(00\)00032-4](https://doi.org/10.1016/S1096-7516(00)00032-4)

Harrison, D. and Rubinfeld, D. 1978. Hedonic prices and the demand for clean air. *Journal of Environmental Economics and Management* 5, 81-102. [https://doi.org/10.1016/0095-0696\(78\)90006-2](https://doi.org/10.1016/0095-0696(78)90006-2)

Helsper, E. and Eynon, R. 2010. Digital natives: where is the evidence? *British Educational Research Journal* 36, 1-18. <https://doi.org/10.1080/01411920902989227>

Hendry, D. 2015. *Introductory macro-econometrics: A new approach*. London: Timberlake Consultants Press.

Hendry, D. 2018. Deciding between alternative approaches in macroeconomics. *International Journal of Forecasting* 34, 119-135. <https://doi.org/10.1016/j.ijforecast.2017.09.003>

Hendry, D. and Mizon, G. 2016. Improving the teaching of econometrics. *Cogent Economics and Finance* 4. <https://doi.org/10.1080/23322039.2016.1170096>

Hendry, D. and Nielsen, B. 2010. **A modern approach to teaching econometrics**. *European Journal of Pure and Applied Mathematics* 3, 347-369.

Hmelo-Silver, C. 2004. Problem-Based Learning: What and how do students learn? *Educational Psychology Review* 16, 235-266.

<https://doi.org/10.1023/B:EDPR.0000034022.16470.f3>

Horton, N. and Hardin, J. 2015. Teaching the next generation of statistics students to “Think With Data”: Special Issue on Statistics and the Undergraduate Curriculum. *The American Statistician* 69, 259-265. <https://doi.org/10.1080/00031305.2015.1094283>

Huang, F., Zheng, S., Fu, P., Tian, Q., Chen, Y., Jiang, Q. and Liao, M. 2023. Distinct classes of statistical anxiety: Latent profile and network psychometrics analysis of university students. *Psychology Research and Behavior Management* 16, 2787-2802  
<https://doi.org/10.2147/PRBM.S417887>

Huang, X. and Mayer, R. 2016. Benefits of adding anxiety-reducing features to a computer-based multimedia lesson on statistics. *Computers in Human Behavior* 63, 293-303.  
<https://doi.org/10.1016/j.chb.2016.05.034>

Huang, X., Mayer, R. and Usher, E. 2020. Better together: Effects of four self-efficacy-building strategies on online statistical learning. *Contemporary Educational Psychology* 63, 101924 <https://doi.org/10.1016/j.cedpsych.2020.101924>

Ioannidis, J. 2005. Why most published research findings are false. *PLoS Med* 2(8): e124.  
<https://doi.org/10.1371/journal.pmed.0020124>

Janz, N. 2016. Bringing the Gold Standard into the classroom: Replication in University teaching. *International Studies Perspectives* 17, 392-407. <https://doi.org/10.1111/insp.12104>

Jerez, O., Orsini, C., Ortiz, C. and Hasbun, B. 2021. Which conditions facilitate the effectiveness of large-group learning activities? A systematic review of research in higher education. *Learning: Research and Practice* 72, 147-164.  
<https://doi.org/10.1080/23735082.2020.1871062>

Jones, C., Ramanau, R., Cross, S. and Healing, G. 2010. Net generation or Digital Natives: Is there a distinct new generation entering university? *Computers and Education* 54, 722-732. <https://doi.org/10.1016/j.compedu.2009.09.022>

Kalyuga, S., Chandler, P., Tuovinen, J., and Sweller, J. 2001. When problem solving is superior to studying worked examples. *Journal of Educational Psychology* 93, 579-588.  
<https://doi.org/10.1037/0022-0663.93.3.579>

Kalyuga, S., Ayres, P., Chandler, P., and Sweller, J. 2003. Expertise reversal effect. *Educational Psychologist* 38, 23-31. [https://doi.org/10.1207/S15326985EP3801\\_4](https://doi.org/10.1207/S15326985EP3801_4)

Kapur, M. 2008. Productive Failure. *Cognition and Instruction* 26, 379-424.  
<https://doi.org/10.1080/07370000802212669>

Kapur, M. 2012. Productive failure in learning the concept of variance. *Instructional Science* 40, 651-672. <https://doi.org/10.1007/s11251-012-9209-6>

Kapur, M. 2015. Learning from productive failure. *Learning: Research and Practice* 1, 51-65.  
<https://doi.org/10.1080/23735082.2015.1002195>

Kirkwood, A. and Price, L. 2013. Technology-enhanced learning and teaching in higher education: what is 'enhanced' and how do we know? A critical literature review. *Learning, Media and Technology* 39, 6-36 <https://doi.org/10.1080/17439884.2013.770404>

Kirschner, P., Sweller, J. and Clark, R. 2006. Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential and inquiry-based teaching. *Educational Psychologist* 41, 75-86. [https://doi.org/10.1207/s15326985ep4102\\_1](https://doi.org/10.1207/s15326985ep4102_1)

Kirschner, P. and van Merriënboer, J. 2013. Do learners really know best? Urban legends in education. *Educational Psychologist* 48, 169-183. <https://doi.org/10.1080/00461520.2013.804395>

Lefrançois, G. 1997. *Psychology for teachers*. 9th ed. Belmont: Wadsworth.

Legacy, C., Le, L., Zieffler, A., Fry, E., and Vivas Corrales, P. 2024. The Teaching of introductory statistics: Results of a national survey. *Journal of Statistics and Data Science Education* 32, 232-240. <https://doi.org/10.1080/26939169.2024.2333732>

Loibl, K., Roll, I. and Rummel, N. 2017. Towards a theory of when and how problem solving followed by instruction supports learning. *Educational Psychology Review* 29, 693-715. <https://doi.org/10.1007/s10648-016-9379-x>

MacInnes, J., Breeze, M., de Haro, M., Kandlik, M. and Karels, M. 2016. **Measuring Up: International Case Studies on the Teaching of Quantitative Methods in the Social Sciences**. London: The British Academy.

Mansell, W. 2015. **Count Us In: Quantitative Skills for a New Generation**. London: British Academy.

Mason, G., Nathan, M. and Rosso, A. 2015. **State of the Nation: A Review of Evidence on the Supply and Demand of Quantitative Skills**. London: British Academy and NIESR.

MacManaway, L. 1970. Teaching methods in higher education- innovation and research. *Universities Quarterly* 24, 321-329. <https://doi.org/10.1111/j.1468-2273.1970.tb00346.x>

Mautone, P. and Mayer, R. 2001. Signaling as a cognitive guide in multimedia learning. *Journal of Educational Psychology*, 932, 377-389. <https://doi.org/10.1037/0022-0663.93.2.377>

Mayer, R. 2004. Should there be a three-strikes rule against pure discovery learning? *American Psychologist* 59, 14-19. <https://doi.org/10.1037/0003-066X.59.1.14>

Mayer, R. 2021. *Multimedia Learning* 3<sup>rd</sup> edition. Cambridge: Cambridge University Press.

Nachtigall, V., Serova, K. and Rummel, N. 2020. When failure fails to be productive: probing the effectiveness of productive failure for learning beyond STEM domains. *Instructional Science* 48, 651-697. <https://doi.org/10.1007/s11251-020-09525-2>

National Academies of Sciences, Engineering and Medicine (NASEM). 2019. **Reproducibility and replicability in science**. Washington, DC: The National Academies Press.

- Neumann, D.L., Hood, M. and Neumann, M. 2013. Using real-life data when teaching statistics: student perceptions of the strategy in an introductory statistics course. *Statistics Education Research Journal* 12, 59-70. <https://doi.org/10.52041/serj.v12i2.304>
- Onwuegbuzie, A. 2004. Academic procrastination and statistics anxiety. *Assessment & Evaluation in Higher Education* 29, 3-19. <https://doi.org/10.1080/0260293042000160384>
- Phillips, D. 1998. How, why, what, when, and where: Perspectives on constructivism in psychology and education. *Issues in Education* 3, 151-194.
- Prensky, M. 2001a. Digital natives, digital immigrants Part 1. *On the Horizon* 9(5), 1-6. <https://doi.org/10.1108/10748120110424816>
- Prensky, M. 2001b. Digital natives, digital immigrants Part 2: Do they really think differently? *On the Horizon* 9(6), 1-6. <https://doi.org/10.1108/10748120110424843>
- Prince, M. 2004. Does active learning work? A review of the research. *Journal of Engineering Education* 93, 223-231. <https://doi.org/10.1002/j.2168-9830.2004.tb00809.x>
- Reade, J. 2007. Modelling and forecasting football attendances. *Oxonomics* 2, 27-32. <https://doi.org/10.1111/j.1752-5209.2007.00015.x>
- Renkl, A. and Atkinson, R. 2003. Structuring the transition from example study to problem solving in cognitive skill acquisition: A cognitive load perspective. *Education Psychologist* 38, 15-22. [https://doi.org/10.1207/S15326985EP3801\\_3](https://doi.org/10.1207/S15326985EP3801_3)
- Roll, I., Holmes, N., Day, J. and Bonn, D. 2012. Evaluating metacognitive scaffolding in guided invention activities. *Instructional Science* 40, 691-710. <https://doi.org/10.1007/s11251-012-9208-7>
- Rosenshine, B. 2012. Principles of instruction: Research-based strategies that all teachers should know. *American Educator* 36, 12-19. [ERIC EJ971753](https://eric.ed.gov/?id=EJ971753)
- Rozgonjuk, D., Kraav, T., Mikkor, K., Orav-Puurand, K. and Täht, K. 2020. Mathematics anxiety among STEM and social sciences students: the roles of mathematics self-efficacy, and deep and surface approach to learning. *International Journal of STEM Education* 7, 46. <https://doi.org/10.1186/s40594-020-00246-z>
- Saunders, F. and Gale, A. 2012. Digital or didactic: Using learning technology to confront the challenge of large cohort teaching. *British Journal of Educational Technology* 43, 847-858. <https://doi.org/10.1111/j.1467-8535.2011.01250.x>
- Savery, J. and Duffy, T. 1995. Problem Based Learning: An instructional model and its constructivist framework. *Educational Technology* 35, 31-38. <https://www.jstor.org/stable/44428296>
- Scott Jones, J. and Goldring, J. 2015. 'I'm not a quants person'; key strategies in building competence and confidence in staff who teach quantitative research methods'. *International Journal of Social Research Methodology* 18, 479-494. <https://doi.org/10.1080/13645579.2015.1062623>



Skulmowski, A. and Xu, K. 2021. Understanding cognitive load in digital and online learning: A new perspective on extraneous cognitive load. *Educational Psychology Review*, 34, 171-196. <https://doi.org/10.1007/s10648-021-09624-7>

Slavin, R. 2006. *Educational psychology: Theory and Practice* (13th ed.). London: Pearson.

Smith, L., Yu, F. and Schmid, K. 2021. Role of replication research in biostatistics graduate education. *Journal of Statistics and Data Science Education*, 291, 95–104. <https://doi.org/10.1080/10691898.2020.1844105>

Spanjers, I., Van Gog, T., Van Merriënboer, J. 2012. Segmentation of Worked Examples: Effects on Cognitive Load and Learning. *Applied Cognitive Psychology* 26, 352-358. <https://doi.org/10.1002/acp.1832>

Stojmenovska D., Bol T. and Leopold T. 2019. Teaching replication to graduate students. *Teaching Sociology*, 474, 303-313. <https://doi.org/10.1177/0092055X19867996>

Sweller, J. and Cooper, G. 1985. The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction* 2, 59-89. [https://doi.org/10.1207/s1532690xci0201\\_3](https://doi.org/10.1207/s1532690xci0201_3)

Sweller, J., Van Merriënboer, J.J.G. and Paas, F. 1998. Cognitive architecture and instructional design. *Educational Psychology Review* 103, 251-296. <https://doi.org/10.1023/A:1022193728205>

Sweller, J., Van Merriënboer, J. and Paas, F. 2019. Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review* 31, 261-292. <https://doi.org/10.1007/s10648-019-09465-5>

Tapscott, D. (1998). *Growing up digital*. San Francisco: McGraw-Hill.

VanLehn, K., Siler, S., Murray, C., Yamauchi, T., and Baggett, W. 2003. Why do only some events cause learning during human tutoring? *Cognition and Instruction* 21, 209-249. [https://doi.org/10.1207/S1532690XCI2103\\_01](https://doi.org/10.1207/S1532690XCI2103_01)

van Merriënboer, J. and Sweller, J. 2005. Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review* 17, 147-177. <https://doi.org/10.1007/s10648-005-3951-0>

Veen, W. and Vrakking, B. 2006. *Homo Zappiens: growing up in a digital age*. London: Network Continuum Education.

Veen, W. and van Staaldouin, J. 2010. The Homo Zappiens and its consequences for learning in Universities. In: Ehlers, U. and Schneckenberg, D. (eds) *Changing Cultures in Higher Education*. Berlin: Springer.

Villarroel, V., Bloxham, S., Bruna, D., Bruna, C., and Herrera-Seda, C. 2017. Authentic assessment: creating a blueprint for course design. *Assessment & Evaluation in Higher Education* 435, 840-854. <https://doi.org/10.1080/02602938.2017.1412396>

Westermann, K., Rummel, N. 2012. Delaying instruction: evidence from a study in a university relearning setting. *Instructional Science* 40, 673-689.

<https://doi.org/10.1007/s11251-012-9207-8>

Williams, J. and Wong, A. 2009. The efficacy of final examinations: A comparative study of closed-book, invigilated exams and open-book, open-web exams. *British Journal of Educational Technology* 40, 227-236. <https://doi.org/10.1111/j.1467-8535.2008.00929.x>

Wooldridge, J. 2020. *Introductory Econometrics: A Modern Approach*. Boston: Cengage.

World Economic Forum. 2019. [Data Science in the New Economy](#).

Zahaciva, A., Lynch, S. and Espenshade, T. 2005. Self-efficacy, stress, and academic success in college. *Research in Higher Education* 46, 677-706. <https://doi.org/10.1007/s11162-004-4139-z>

## Notes

[1] Discussion of the importance of data literacy and data science skills, the demand for graduates with statistical skills and a need for universities to address a lack of data literacy skills extend beyond the Social Sciences. See, inter alia, [Coners et al. \(2024\)](#), [Horton and Harin \(2015\)](#) and [Ghodoosi et al. \(2023\)](#).

[2] As [Buchele \(2020\)](#) emphasise engagement and attendance are not the same.

[3] The econometrics software employed for analysis in this chapter is EViews 13. Given the nature of the data considered here, the Durbin-Watson statistic automatically produced in the output is removed from the tables presented in this paper.

[4] In this example, specific elements have been redacted from Table Three. However, additional flexibility can be introduced by varying which elements are removed. Adjusting the redacted components allows for a greater challenge, as students must identify which missing elements are relevant to the task at hand. For example, if only the F-statistic and its p-value were removed, their absence would emphasise their importance, potentially simplifying the process of answering the question.

[5] Data scaling is a familiar issue when considering linear regression. See, for example, [Wooldridge \(2020, section 6.1\)](#).

[6] The notion of impasse-driven learning (see, inter alia, [VanLehn et al., 2003](#)) can be related to the concept of productive failure.

[7] An excellent illustration and discussion of the complexities exhibited by real-world economic data as a result of economic and political events is provided by [Hendry and Nielsen \(2010\)](#).

Views on request

©2025 UNIVERSITY OF BRISTOL  
EXCEPT WHERE STATED

