Pearson

Efficacy Reporting Framework

March 20, 2019

Introduction

Pearson's mission is to help people make progress in their lives through learning. But helping people to achieve the learning outcomes that matter most in life, like new knowledge and skills to support progression into further or higher education, or in a career, isn't something that happens by accident. It happens by *design*.

When we first start our learning journey, the choices of parents and educators often drive decisions about learning. As we grow, we take over from parents and educators, becoming the designers of our own lifelong learning journey. We identify the outcomes we want to achieve the most, and select learning experiences that suit how we want to access and engage with learning.

At Pearson, we are committed to supporting you to achieve the outcomes that matter most to *you*. That's why we design products focused on supporting the achievement of those outcomes, why we underpin the design and implementation support with evidence about what works to improve teaching and learning, and why we measure the impact of use of our products on outcomes. We use what we learn to continuously improve how our products and services are designed and used.

This is what our commitment to efficacy means, and why it is at the core of our mission as a company. It brings rigor to how we break down big learning goals into manageable steps, and focus and alignment to how we combine our world leading content, assessment, and technology capabilities into learning experiences.

We also publicly report on the efficacy of use of our products. Our reports help us, and the wider education community, build a better understanding of not just what works, but how, why, and in what context — helping us learn, not guess, about how the design and use of products relates to the achievement of outcomes that matter most to you.

The 2019 Product Efficacy Reports include three audited, standards-based efficacy research studies: on Revel for Psychology, 1st edition by Marin and Hock in North America, MyPedia in India and Sistema COC in Brazil. We are simultaneously publishing non-audited efficacy reports on two of our most frequently used assessment and qualifications products – Pearson Test of English Academic and the UK regulated GCSE Mathematics qualification.

What's new in 2019

Following the release of the 2018 efficacy reports, we gathered feedback from educators and thought leaders around the world. Our commitment to open, transparent efficacy reporting was received positively, but we were encouraged to do even more to provide support about how teaching and learning practices can, and do, change when using digital products. We are responding in 2019 by providing even more details about the specific implementations featured in the impact evaluation research. Starting with Revel for *Psychology*, 1st edition by Marin and Hock, we are releasing companion educator guides alongside each report, highlighting the educators and students behind the outcomes.

We remain committed to continuously improving how we are applying efficacy in education, all with a focus on helping more people make progress in their lives through learning. This sense of purpose gives us a reason to keep on fighting, nothing spared, to improve how we do things in education.

Kate Edwards
Senior Vice President,
Efficacy and Learning Research, Pearson
March 20, 2019

Special thanks

We want to thank all the educators, students, research institutions, and organizations who have collaborated with us. Of course, our work isn't possible without partnerships within Pearson and the support of our leadership. We also want to thank PwC and SRI Education for their contribution to our rigorous standards and quality.

If you are interested in partnering with us on future efficacy research, have feedback or suggestions for how we can improve, or want to discuss your approach to using or researching our products, we would love to hear from you at **efficacy@pearson.com**. If we, as a sector, tackle this together, we will help more learners, learn more.

Contents

About this document

This document explains how Pearson approaches efficacy reporting.

Part 1 explains what we hope to achieve by our efficacy reporting and how we have designed our process to achieve this. Anyone interested in evaluating our process and the rigor behind our findings should find this section useful.

Part 2 goes into detail about the activities we perform and the documentation we use at every stage of the efficacy reporting process. This section also includes the complete guidance we follow when preparing evidence-based efficacy statements, and a breakdown of the work carried out by our auditors. Anyone who is thinking of following a similar process to prepare their own efficacy reporting, or who is simply in search of more granular detail than we provide in Part 1, should find this section useful.

We refer to our approach to efficacy reporting — including the process, controls, documentation (such as our guidance on efficacy statement terminology by research design type), reviews and third-party audit — as the Pearson Efficacy Reporting Framework.

We are making the details of the Efficacy Reporting Framework public in the interests of transparency, and in the hope that sharing our approach will encourage others to give feedback about how we can improve it, or replicate and build on what we have done. This document is published under a Creative Commons licence, so you may share and use it freely as long as you reference it in any work you produce as a result. If you do decide to follow our process, and you want to discuss any aspect of it with us, please contact **efficacy@pearson.com.**

Part 1: Overview of the Efficacy Reporting Framework

Background to efficacy reporting

Why we publish efficacy reporting

At Pearson, we recognize how important it is to understand the relationships between the design and use of our products and the outcomes that matter most to students and educators. We want to understand, not only what works, but also for whom and why.

We engage in research on our products to improve our understanding in these areas. Our focus on improving the efficacy of our products, by applying evidence-based insight both from learning research and design and from impact evaluation research, allows us to relentlessly improve our products so they can support delivery of better outcomes for more learners.

In 2013, we made a public commitment to begin regularly publishing the results of this research by 2018. Rigorous impact evaluation research and transparent reporting allow us to help more learners, learn more by:

- Better understanding and highlighting relationships between the use of our products and the learning and learner outcomes that matter most to students and educators, and using this understanding to help others achieve comparable or even better impact themselves
- Sharing the evidence underpinning the design and development of new and existing products
- Galvanizing other learning companies to follow suit and measure their impact by the outcomes they deliver for learners

Our journey

Year	Action
2013	Pearson commits to start publishing annual reports about the efficacy of its products, and to have those reports externally audited by 2018. Pearson publishes **The incomplete guide to delivering learning outcomes** with the Shared Value Initiative. Pearson publishes **From good intentions to real impact**, an initial approach to efficacy research and reporting developed in partnership with Nesta.
2014	Pearson appoints PricewaterhouseCoopers LLP (PwC) as its independent auditor.
2014–2016	Pearson refines its approach with input from organizations like the American Educational Research Association, What Works Clearinghouse, and the Efficacy Academic Network.
2015	The Efficacy Academic Network is formed. It is made up of four leading academics from the US, UK, and Australia: Eva Baker, Philippa Cordingley, Chris Dede, and Gordon Stanley. The first dry run of the efficacy reporting process: Pearson publishes reports on five of its products.
2016	The second dry run of the efficacy reporting process: Pearson goes deeper and broader with efficacy reporting — looking at more products and detailing more rigorous research.
2017	The third dry run of the efficacy reporting process: Pearson publishes five reports and this time also subjects them to a mock audit. Pearson examines its approach in more detail with SRI Education, with the results published as **Understand, implement, evaluate.** Pearson sponsors and contributes to the EdTech Efficacy Research Academic Symposium, fostering and contributing to discussion and debate around efficacy.
2018	Pearson publishes its Efficacy Reporting Framework and transparent, rigorous reports about the efficacy of its most widely used digital products. The Product Efficacy Reports were independently assured by PwC.
2019	Pearson uses feedback from 2018 to improve its approach to efficacy reporting — streamlining the process, formalizing and strengthening the role of independent peer reviewers, and publishing Educator Guides alongside the efficacy reports, to provide additional insight and support for implementation. Building on its 2013 commitment, Pearson publishes three new independently assured reports about the efficacy of its products.

What efficacy reporting looks like

Key findings from Pearson's research are reported in the form of efficacy statements. These are evidence-based statements summarizing the primary conclusions of a study. Each statement relates to the use of a Pearson product and/or its impact on learners and learning.

Each efficacy statement represents a finding from a specific study using a particular type of research design. The effects described in an efficacy statement are dependent on the research design used and its limitations. For that reason, they should be interpreted within the context of the research design used that supports the evidence and details how, when, and where the evidence was collected.

To be as open and transparent as possible about how we design, develop, and evaluate the impact of use of our products on learning, Pearson publishes two different kinds of efficacy reports. Each kind of report provides a different level of insight into the research supporting the efficacy statements relating to the use of the product.

Technical Research Reports each describe a single piece of research into the use of a product, undertaken or commissioned by Pearson to meet the standards expected for publication in peer-reviewed academic journals. Technical Research Reports include enough detail for another researcher to replicate the research.

Product Efficacy Reports summarize all the relevant research related to the use of a single product, based on one or more Pearson Technical Research Reports, as well as any external studies we discover that meet our criteria for inclusion. They also include information about the learning research that informed the product's design, and how the product is designed to be used.

Types of research design for efficacy reporting

The research that underpins our audited efficacy statements includes correlational designs, comparative designs, and causal designs.

- **Correlational designs** study students using Pearson products to investigate whether there is a relationship between how they use the product and its intended outcomes for learners, while controlling for background factors like baseline achievement.
- **Comparative designs** compare learner outcomes for students who use a Pearson product to those of students who don't, but those student groups may be very different from one another, which means we cannot rule out confounding factors.
- **Causal designs** compare learner outcomes for students who use a Pearson product to those of students who don't, with student groups that are highly similar at baseline, allowing us to rule out as many confounding factors as possible.

It should be noted that Pearson makes four types of efficacy statements derived from these three types of research design, as explained on page 18. In addition to correlational, comparative, and causal efficacy statements, we also make descriptive efficacy statements, which can be derived from any study capable of supporting another type of efficacy statement. Descriptive efficacy statements are purely statements about the sample, not any wider population.

Pearson makes use of a wider range of research activity types than this throughout the lifecycle of a product, including market research and exploratory learning research. But when we prepare efficacy reporting and audited efficacy statements for a product, we use these three specific types of study design.

This selection of study design types for our research allows us to build a body of evidence over time, and to use the evidence to demonstrate impact to our customers, improve existing products, and develop new ones.

Reviewing and auditing efficacy reports

It is vitally important that our customers know they can trust the statements we make about the use of our products.

A team of researchers from SRI Education with expertise in study design and quantitative methodology performed an independent review of Pearson's research and efficacy statements relating to the use of the product and its impact on learners and learning.

First, the team helped Pearson codify its research quality criteria into tiered levels of evidence to reflect best practices in the field. Criteria for these tiered levels of evidence drew both from the tiered evidence levels in the USA's Every Student Succeeds Act (ESSA) and the 2006 American Educational Research Association's Standards for Reporting on Empirical Social Science Research.

Then the team assessed whether Pearson's analyses, technical reporting, and associated efficacy statements complied with these research quality criteria. This year's reviews applied Pearson's comparative levels of evidence, which draw from ESSA's "promising" level of evidence, and correlational level of evidence, which adopt the standards for studies that do not include a comparison group.

We have appointed PricewaterhouseCoopers LLP (PwC) to provide limited assurance on the efficacy statements set out in our Product Efficacy Reports. Each Product Efficacy Report includes an independent assurance report from PwC on the efficacy statements about the product. This is to demonstrate that the statements accurately reflect the research that has been carried out. As well as the efficacy statements themselves, PwC's audit involves evaluating the design and operation of the process we use to generate the statements.

The efficacy reporting process

Pearson has developed a four-stage process to prepare reporting on the impact of use of our products on outcomes.

Process 1: Learner outcomes
In partnership with our customers and learners, we define the outcomes that matter to learners and the metrics for measuring those outcomes.

Process 2: Study preparation
We conduct a literature review to discover any existing studies that meet our relevance and rigor criteria, and design our own studies, to gather evidence about how the use of our products is related to learner outcomes.

Process 3: Study write-up and draft efficacy statements
The prepared impact evaluation research is conducted, either by Pearson or by a commissioned researcher. This is independently reviewed by SRI Education to make sure that our findings and statements are accurately backed up by evidence. Pearson also assesses the rigor and completeness of the data that supports the efficacy statements.

Process 4: Product Efficacy Reports
If the literature review in process 2 was conducted more than 12 months before the reporting date, we conduct a second literature review to discover any recent studies that meet our relevance and rigor criteria. We then synthesize all findings — including evidence from research we have conducted or commissioned, and evidence from pre-existing research — and present them in an appropriate manner. We also include information about the learning research that underpins the design of the product, although this part is not subject to audit.

The process includes controls designed to make sure our reporting of the research we conduct is rigorous and accurate. Each stage of the process also has a set of associated review documents, or checklists. These documents:

- ensure that key activities are applied consistently across all products
- form a record of the key activities completed
- inform improvements to the product in the future

The prominent academics of the Efficacy Academic Network provided feedback during the development of the efficacy reporting process. This feedback, the independent assurance from PwC, and the independent review by SRI Education are all ways in which we improve the rigor of our efficacy reporting.

Creating accurate efficacy statements

We need our customers to know they can trust what we tell them about our products. It is vitally important, therefore, that our efficacy statements accurately represent the evidence from the research they are based on.

This means the wording of efficacy statements must be carefully considered to make sure it is aligned with the standard and design of the research it is derived from. We make four types of efficacy statements. Three of these map directly to the three research design types.

- **Descriptive** efficacy statements can correspond to correlational, comparative, or causal designs.
- **Correlational** efficacy statements correspond to correlational designs.
- **Comparative** efficacy statements correspond to comparative designs.
- **Causal** efficacy statements correspond to causal designs.

For more detail on the different types of research design, specific research standards applicable to each research design, and example efficacy statements, please see page 18.

Part 2: The Efficacy Reporting Framework in detail

Process 1: Learner outcomes

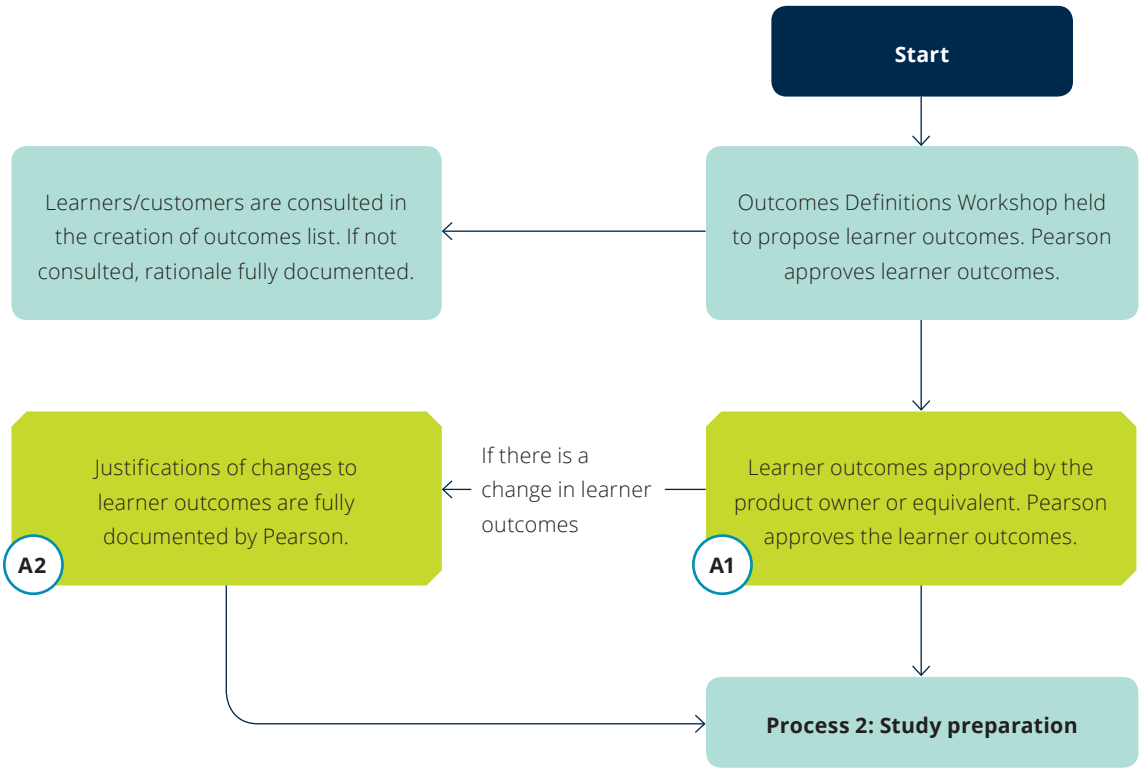
In Process 1, in partnership with customers and learners, Pearson defines the outcomes that matter to learners and the metrics for measuring those outcomes.

Artefacts required in process 1

Reference	Artefact	Purpose
A1	Learner outcomes form	Documents the agreed intended learner outcomes for a product, as well as the metrics and leading indicators that can be used to measure them
A2	Rationale for all changes to learner outcomes	Documents changes made to the product's learner outcomes over time

What's involved

Process 1: Learner outcomes



Key

- Relevant process for audit
- Relevant control for audit
- (A) Relevant artefact for audit

Process 2: Study preparation

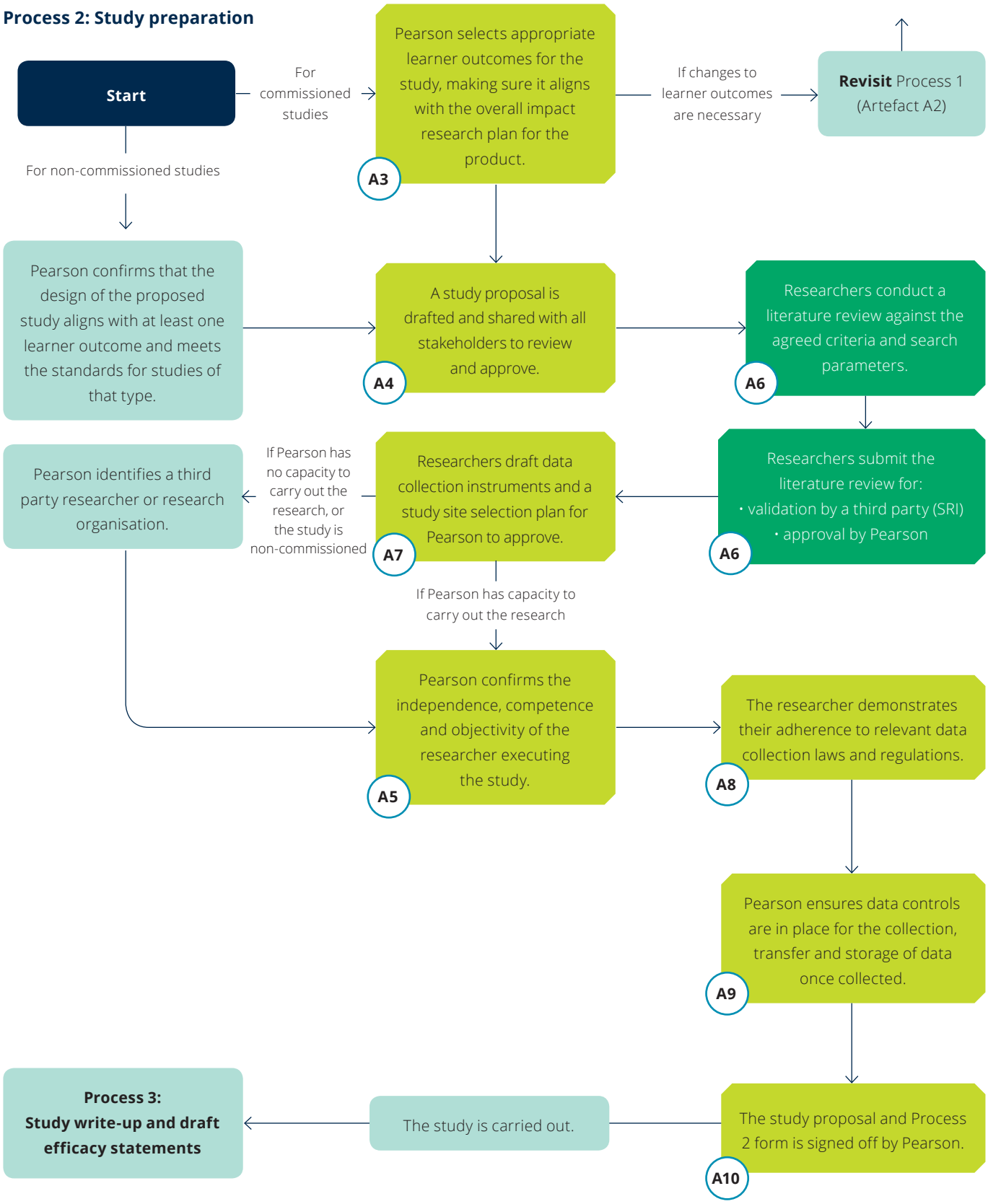
In Process 2, we conduct a literature review to discover any existing studies that meet our relevance and rigor criteria, and design our own studies, to gather evidence about how the use of our products is related to learner outcomes.

Artefacts required in process 2

Reference	Artefact	Purpose
A3	Impact evaluation research plan for the product	Records all planned impact evaluation research to be conducted for a product
A4	Study proposal	Details a proposed impact evaluation research study
A5	Independence, competence, and objectivity confirmation form	Provides evidence that the lead researcher is able to conduct the research study objectively
A6	Literature review tracker	Provides evidence of the literature search methods and results in terms of relevance and adherence to research quality standards of the impact evaluation research studies found
A7	Draft data collection instruments	Document any data collection instruments proposed for use in the study
A8	Evidence of compliance before data collection	Provides evidence that the researcher has complied with any data collection rules, or sought relevant data use permissions, before the study is conducted (for example, a Data Use Agreement or evidence of approval or exempt status from an Institutional Review Board or other ethics board)
A9	Evidence of compliance following data collection	Provides evidence to confirm that the researcher complied with any data collection rules stipulated in the Data Use Agreement or Institutional Review Board/ethics board documentation during the process of conducting the study, and that appropriate controls were in place for the collection, transfer, and storage of data
A10	Process 2 form	Documents sign-off from Pearson team members that all relevant evidence in Process 2 has been collected and correctly documented

What's involved

Process 2: Study preparation



Key

- Relevant process for audit
- Relevant control for audit
- Control not relevant for audit
- (A) Relevant artefact for audit

Process 3: Study write-up and draft efficacy statements

In Process 3, the prepared impact evaluation research is conducted, either by Pearson or by a commissioned researcher. This is independently reviewed by SRI Education to make sure that our findings and statements are accurately backed up by evidence. Pearson also assesses the rigor and completeness of the data that supports the efficacy statements.

Artefacts required at process 3

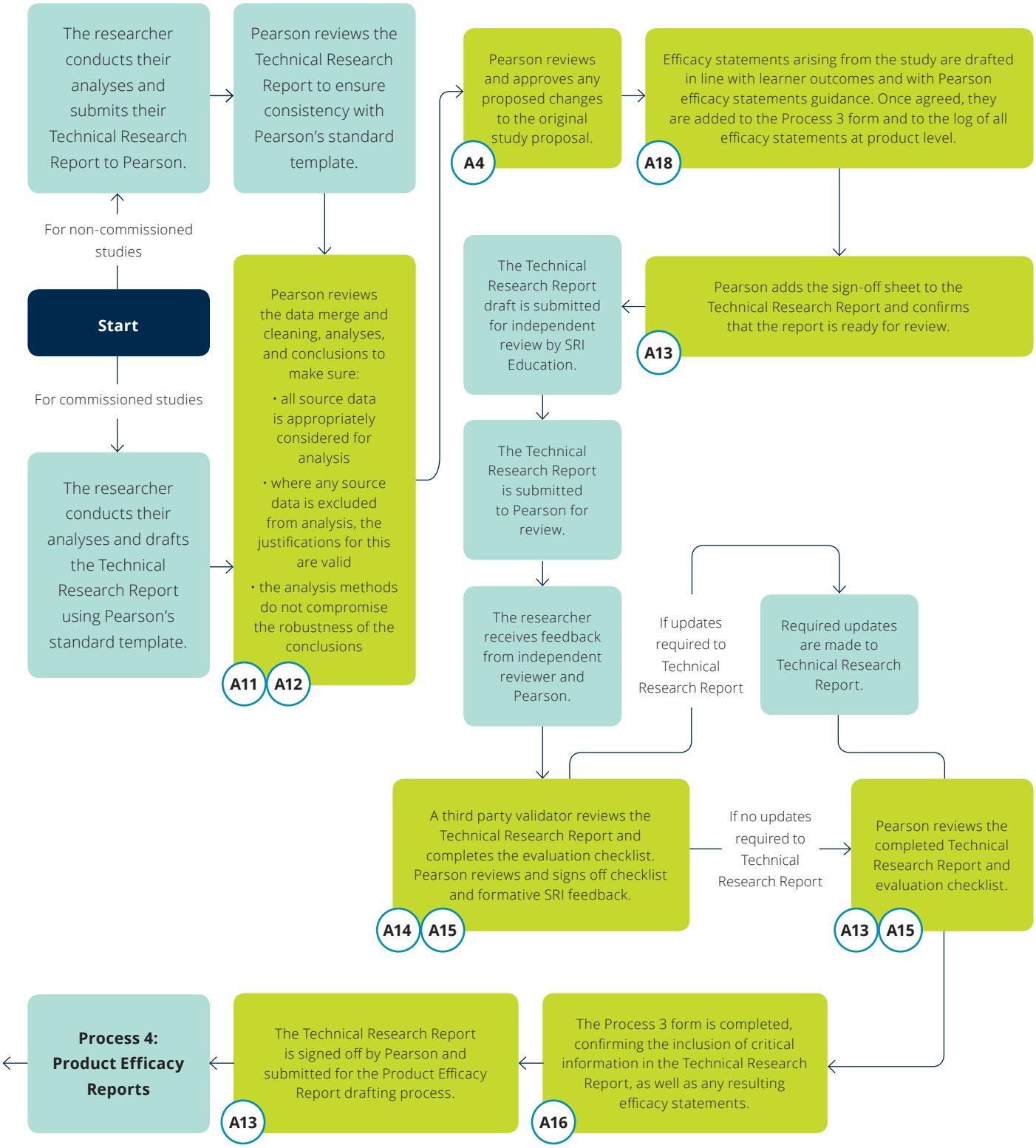
Reference	Artefact	Purpose
A11	Statistical output used to populate the results tables	Documents any statistical codes and outputs used to generate statistical results and tables
A12	Data merge and clean form	Provides evidence that a senior member of Pearson research staff has checked the way the researcher merged and cleaned the data to ensure accuracy and fidelity
A13	Technical Research Report	Documents the study, including methodology, results, efficacy statements, and any limitations, in enough detail for another researcher to replicate the study
A14	Write-up from independent reviewer	Presents formative feedback from an independent third party (other than PwC), which the Pearson researcher either acts on by updating the Technical Research Report or responds to with a rationale for why the changes were not made
A15	Technical Research Report review checklist	Documents sign-off by SRI Education that the Technical Research Report meets Pearson's standards for conducting a study, and that any efficacy statements are supported by the research
A16	Process 3 form	Documents sign-off from Pearson team members that all relevant evidence in Process 3 has been collected and correctly documented

Artefact resubmissions

Reference	Artefact	Purpose
A4	Study proposal (updated)	Provides evidence and a rationale for any changes that were made to the original study proposal in the course of conducting the study due to unforeseen events (e.g., attrition, missing data, treatment contamination)

What's involved

Process 3: Study write-up and draft efficacy statements



Key

- Relevant process for audit
- Relevant control for audit
- (A) Relevant artefact for audit

Process 4: Product Efficacy Reports

In Process 4, we synthesize all findings from research we have conducted or commissioned, and evidence from pre-existing research — and present them in an appropriate manner. We also include information about the learning research that underpins the design of the product, although this part is not subject to audit.

If the literature review in Process 2 was conducted more than 12 months before the reporting date, we also conduct a second literature review in Process 4, to discover any recent studies that meet our relevance and rigor criteria.

Artefacts required in process 4

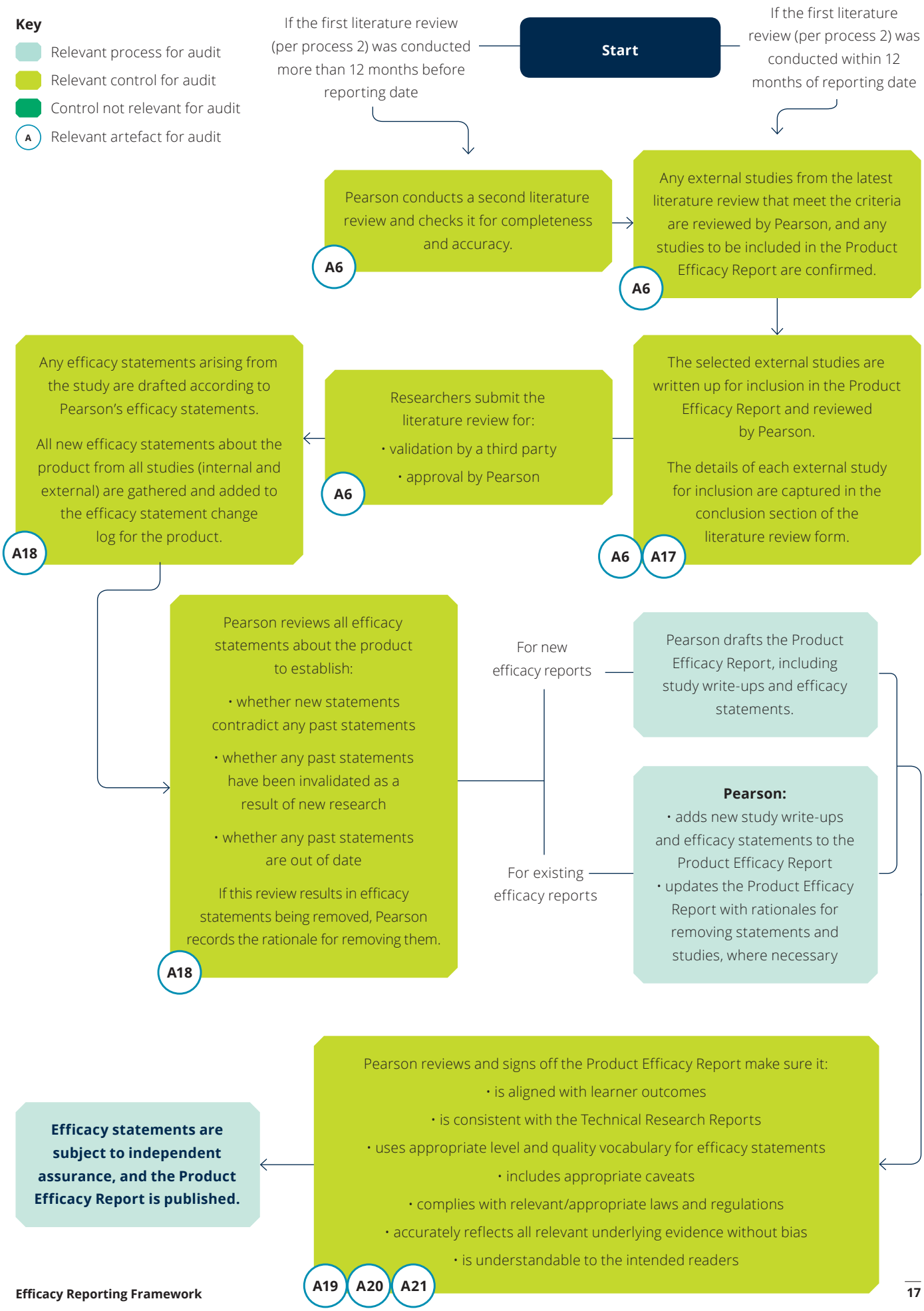
Reference	Artefact	Purpose
A17	Literature review write-up	Documents the methods and results of the literature review, for inclusion in the Product Efficacy Report
A18	Log of all efficacy statements related to the product	Provides evidence that all efficacy statements have been agreed and signed off by the relevant stakeholders Provides a rationale for any changes to be made to previously published efficacy statements
A19	Draft Product Efficacy Report	Provides stakeholders (including PwC) with an opportunity to review and feed back on the content of the Product Efficacy Report that summarizes all relevant impact evaluation research related to the use of a product, as well as any external studies found that meet the criteria for inclusion
A20	Final Product Efficacy Report	Summarizes all the relevant impact evaluation research related to the use of a product, as well as any external studies found that meet the criteria for inclusion
A21	Process 4 form	Documents sign-off from senior Pearson team members, including the senior vice president of efficacy and research, that all relevant evidence in Process 4 has been collected and correctly documented

Artefact resubmissions

Reference	Resubmitted artefact	Purpose
A6	Literature review tracker (updated)	Documents any additional research that has been published since the last literature review was conducted and is eligible for inclusion in the Product Efficacy Report. If the literature review was conducted more than 12 months before the reporting date, a second literature review is performed, and the results recorded in this tracker

What's involved

Process 4: Product Efficacy Reports



Efficacy statement terminology by research design type

Pearson uses the following guidance when preparing efficacy statements about the use of its products. The aim of the guidance is to ensure that efficacy statements accurately represent the evidence they are based on, by aligning their wording with the standard and design of the underlying studies.

The standards by which Pearson judges studies are informed by and aligned with recognized research standards and/or frameworks developed by the What Works Clearinghouse and the American Education Research Association.

The illustrative example efficacy statements presented here should not be construed as comprehensive or exhaustive. Rather, they are intended to illustrate the types of wording that would be acceptable for a given combination of efficacy statement type and underlying evidence. The wording of efficacy statements may vary slightly, while still preserving their meaning.

Where a study investigated a relationship or difference between factors, Pearson will include efficacy statements that describe the relationship or difference found. Where the study found that the relationship or difference was not statistically significant, the reports will state this clearly.

Types of efficacy statement

Efficacy statements can be categorized as either general or specific. Alongside this broad categorization, efficacy statements may be:

- descriptive
- correlational
- comparative
- causal

General versus specific efficacy statements

General efficacy statements express a trend, relationship, or effect in simple and broad terms.

Specific efficacy statements provide more detail by:

- communicating more precise numeric estimates
- representing effect sizes or mean differences along with relevant statistical tests
- interpreting general statements in terms of practical significance

Typically, efficacy reporting will use both general and specific efficacy statements.

Requirements of specific efficacy statements

To be considered complete and accurate, specific efficacy statements must include certain important components.

Descriptive efficacy statements

Which research designs support descriptive statements?

Any research design capable of supporting correlational, comparative, or causal efficacy statements can also support descriptive statements via analysis of surveys, secondary analysis of administrative data, or cohort analysis with no controls for baseline factors or potential mediating/moderator variables.

What are the essential requirements of descriptive statements?

- Result phrased as a specific quantified measure (such as a percentage or student count) supported by the study's analysis
- Refers to the specific study sample to which the statement applies (for example, all respondents to a survey, female students using the product/service, grade levels)
- Refers to a learner outcome of interest
- Language is descriptive in nature and does not express a quantitative correlation or statistical association

Illustrative examples of descriptive statements

- Students generally felt the product helped them achieve their goal
- #% of students indicate the product helps them improve their learning "significantly" or "very significantly"

Correlational efficacy statements

Which research designs support correlational statements?

Cohort analysis with controls for baseline factors or potential mediating/moderator variables.

What are the essential requirements of correlational statements?

- Based on a specific, quantified relationship between one malleable measurement (like learner behavior or product/service usage) and one learner outcome of interest
- Refers to the specific study sample to which the statement applies
- Language expresses a directional correlation or statistical association

Illustrative examples of correlational statements

- Among students using the product, learners who accessed more hints were more likely to earn a final grade of B or better
- Learners completing # additional homework problems were # times more likely to pass the course

Comparative efficacy statements

Which research designs support comparative statements?

Those that use a treatment and comparison group, with controls for baseline factors, although baseline equivalence of the two groups is not required. Designs may include quasi-experimental studies or randomized controlled trials where baseline equivalence is not achieved or where attrition patterns do not meet What Works Clearinghouse guidelines.

What are the essential requirements of comparative statements?

- Contains a specific, quantified measure of size (such as coefficient or effect size), significance, and direction of group difference based on analysis results described in the study
- Refers to the applicable treatment and comparison groups
- Based on measurements from a randomized controlled trial or quasi-experimental design with:
  - pre-/post-measurements (the pre-test measure must be a measure of achievement and must be correlated with the outcome measure, but need not be in the same domain)
  - a comparison group that was not exposed to the intervention
  - (for a randomized controlled trial) a comparison group that was determined at random
- Based on a design that avoids n=1 perfect confounding
- Based on analysis that describes the extent to which the groups were similar or different and matched (if appropriate) on key baseline characteristics
- Based on analysis statistically controlling for key characteristics that are greater than 0.25 standard deviations different at baseline
- Language expresses statistical associations

Illustrative examples of comparative statements

- Learners using the product reported a better learning experience than learners who did not use the product, when controlling for prior achievement
- Learners using the product were #% more likely to progress to college level than learners who used a competing product, when controlling for prior achievement

Causal efficacy statements

Which research designs support causal statements?

Those that use a treatment and comparison group, and demonstrate baseline equivalence of the two groups. Attrition must be within What Works Clearinghouse guidelines. Designs may include quasi-experimental approaches (propensity score matching, instrumental variables, regression discontinuity, fuzzy regression discontinuity) or randomized controlled trials.

What are the essential requirements of causal statements?

- Contains a specific, quantified measure of size (such as coefficient or effect size), significance, and direction of group difference based on analysis results described in the study
- Refers to the applicable treatment and comparison groups
- Based on measurements from a randomized controlled trial or quasi-experimental design with:
  - pre-/post-measurements
  - a comparison group that was not exposed to the intervention
  - (for a randomized controlled trial) a comparison group that was determined at random
- Based on a design that avoids:
  - n=1 perfect confounding
  - systematic differences between treatment and control on some observed covariate related to the outcome (such as if the treatment and control outcomes were measured in different cohorts or states)
  - confounding with a second intervention, unless the study intends to capture their joint efficacy (for example, if all treatment classrooms receive both MyLab Statistics and MyLab Mathematics, the treatment effect must be described as the effect of both products)

If using a randomized controlled trial design:

- All assignments determined at random
- Method of randomization explicitly described (for example, student-level randomization or school-level randomization)
- Analytic sample excludes late joiners
- Reports on overall and differential attrition at student and, if applicable, cluster level
- Overall and differential attrition are compared to What Works Clearinghouse bounds (WWC Standards Handbook v.4 p.9–14) and authors acknowledge the potential for bias relative to WWC conservative and liberal attrition standards
- If attrition does not meet the conservative boundary, meets quasi-experimental design baseline equivalence standard

If using a quasi-experimental design:

- Baseline equivalence was achieved on prior achievement according to WWC guidelines (the baseline difference between treatment and control is within .25 SD and, if greater than .05, the measure is included as a covariate)
- Baseline equivalence was achieved for any other variables identified by the study as key characteristics (for example, student mobility, if student mobility is highly relevant to the context of the study)
- Other key confounding variables identified are included as covariates in the analysis
- Language refers to causal effects due to the treatment

Illustrative examples of causal statements

- Learners using the product earned significantly higher final exam scores than similar learners using a competitor product
- Learners using the product achieve #% higher on course tests compared to similar learners who did not use the product

To see the guidance document Pearson refers to when creating efficacy statements, see **appendix 1.**

Research design standards

Pearson developed a set of research quality criteria for guiding research design, and worked with SRI Education to codify these research quality criteria into tiered levels of evidence to reflect best practices in the field. Criteria for these tiered levels of evidence drew both from the tiered evidence levels in the USA's Every Student Succeeds Act (ESSA) and the 2006 American Educational Research Association's Standards for Reporting on Empirical Social Science Research.

Before deriving efficacy statements from any type of study, we review the research to make sure it meets these research quality criteria.

Review category	Review questions
Main findings	Does the research clearly identify "main findings" for the reader? Did the main findings include an analysis of the full sample of interest? (This full sample may exclude students because of missing data.) Were overall outcome measures used in the main findings (for example, using a full assessment score and not using sub-scales)? Were the relevant outcomes for the main findings measured at one or more specific times appropriate for the study's research questions? For studies with multiple main findings in the same domain or with main findings with multiple comparison or treatment groups, were appropriate multiple comparisons corrections consistently applied? (Benjamini-Hochberg correction is preferred.)
Outcome measures	Are the outcome measures used in the main findings of the study aligned to the identified outcomes for the product? For example: • Achievement: assessment measuring students' academic achievement • Timeliness and completion: students' successful completion of, and/or on-time progression through, courses or grade levels • Progression: students' successful completion of, and/or progression through, courses, grade levels, graduation rates, accessibility, and similar types of administrative measures • Employment, earnings, and other outcomes that are relevant to students' educational goals • Social and behavioral competencies, skills, attitudes, and behaviors, including engagement and self-efficacy • Other outcomes that are believed to have positive effects on the above outcomes (for example, instructors' self-efficacy can affect student outcomes, and cost effectiveness can improve student access) Do the outcome measures used in the main findings have face validity? That is, would experts and stakeholders agree that the outcomes appear to measure the constructs well? Are the outcome measures used in the main findings not overly aligned with the intervention? If the outcome measures are within Pearson's control (for example, a Pearson assessment, usage of a Pearson product), do the academic outcome measures used in the main findings meet the What Works Clearinghouse reliability standards for at least one of the following? • internal consistency of 0.50 or higher • temporal stability/test–retest reliability of 0.40 or higher • inter-rater reliability (such as percentage agreement, correlation, or kappa) of 0.50 or higher Do the main findings include analysis of at least one of the following academic outcome measures using the full study population? • an external/third-party test with established validity and reliability evidence • a widely used standardized assessment in a relevant domain (such as a state achievement test or college placement exam) • instructor-made course exams • course grades

	• Pearson's internal, summative platform measures — in which case it must meet:
	  • internal consistency of 0.50 or higher
	  • documentation of content-related validity evidence
	  • items used on the outcome measure must be new to the students (in other words, test–retest validity threat must be controlled for)
	For studies with multiple conditions, were outcome measures for main effects collected in the same manner for all conditions?

Research study sample	Are the overall study sample and analytic samples relevant to the target population of the study (for example, do they draw from one of the key market segments identified for a Pearson product)? Do student-based analyses have analytic sample sizes of at least 100 students? Do teacher-based analyses include at least 66% of the relevant teachers? (For example, if the study includes a survey of teachers, was the survey response rate at least 66%?)
Handling of missing data	Assuming missing data are missing completely at random or missing at random, are missing data consistently handled using one or more of the following appropriate methods? • complete case analysis • regression imputation • full information maximum likelihood • non-response weighting • replacing missing data with a constant, combined with a missing data indicator • another approach justified as appropriate to the case and demonstrated within the study
Handling of nested data	Do the non-descriptive analyses need to account for nested data? Are nested data consistently handled using one or more of the following appropriate methods? • intraclass correlation coefficient calculated to be 0.05 or less and nesting therefore not modeled • multi-level modeling • clusters modeled using fixed effects (applied only if treatment varies within cluster) • clustering correction applied to standard errors • another approach justified as appropriate to the case and demonstrated within the study
Covariates	Do the non-descriptive student analyses include at least one exogenous student-level prior achievement covariate, which must be: • exogenous, that is, collected in such a way that student scores are independent of the study (such as selection into treatment or the effects of the treatment itself) • described to the reader along with its relevant domain (mathematics achievement, for example) and timing of collection • measured at baseline (including shortly after treatment begins) • reflective of achievement (such as cumulative GPA, standardized scores, pre-test performance, or diagnostic test score)

Audit of efficacy statements

Pearson has commissioned PwC to audit the efficacy statements identified in our Product Efficacy Reports. The PwC assurance report is included within each Product Efficacy Report.

The scope, approach and limitations of PwC's work are set out below.

Professional standards applied and level of assurance

PwC performs a limited assurance engagement in accordance with International Standard on Assurance Engagements 3000 (Revised), *Assurance Engagements Other Than Audits and Reviews of Historical Financial Information,* issued by the International Auditing and Assurance Standards board.

A limited assurance engagement is substantially smaller in scope than a reasonable assurance engagement in relation to both the risk assessment procedures (including an understanding of internal control) and the procedures performed in response to the assessed risks.

What is a material misstatement?

A material misstatement would be an efficacy statement that does not reflect the study design and quality of underlying research or the omission of key information from a relevant study.

Work performed by PwC

PwC's audit focuses on:

- the process that generates the efficacy statements that appear in the Product Efficacy Report
- the integrity of the efficacy statements themselves

PwC's work includes the following procedures:

- Making enquiries of relevant Pearson management
- Evaluating the design of the Efficacy Reporting Framework including key structures, systems, processes and controls for managing, generating and reporting the efficacy statements
- Testing all the controls across the four processes of the Efficacy Reporting Framework
- Confirming that all management reviews are performed by at least two members of Pearson's efficacy and research team
- Performing substantive testing, on a sample basis, of the data that underpins the research studies and the resulting efficacy statements, and the controls over the completeness and accuracy of that data (supported by Pearson's internal audit team in those instances where student data is subject to confidentiality restrictions)
- Assessing the quality and conclusions of the underlying research studies
- Inspecting the statistical analysis to assess whether the efficacy statements are valid, supportable and consistent with the underlying research studies
- Independently re-performing the screening of relevant external public research studies and comparing to that done by Pearson
- Assessing the efficacy statements and underlying Technical Research Report(s) for consistency with the Efficacy Reporting Framework
- Reviewing the Product Efficacy Report and Technical Research Report(s) for alignment of research studies and efficacy statements

Important limitations

- The assurance report is product-specific.
- Efficacy research studies reflect the implementation and use of a product in a particular context. It would not be appropriate to assume a product would always generate similar outcomes in the future.

Appendix 1: Pearson efficacy statements guidance

Necessary components for specific statements by statement type:

Typically, both general and specific statements will be used in a report. General statements express a trend, relationship, or effect in simple and broad terms. Specific statements provide more detail by:

- Communicating more precise numeric estimates,
- Representing effect sizes or mean differences along with statistical test statistics,
- Interpreting general statements in terms of practical significance

Specific statements should include important components to be considered complete and accurate.

Note: Example statements should not be construed as comprehensive or exhaustive. Rather, they are intended to serve as illustrative examples of the types of wording that would be acceptable for a given type of statement and evidence. There may be slight variations in wording of the statements, while still preserving meaning. We will only interpret differences/ relationships that are significant. For any non-significant findings, we will clearly state that there are no significant differences or no significant relationship.

Descriptive	Efficacy statement category	Illustrative example statement	Research design used by Pearson associated with statement type	Analysis appropriate for statement *Note: this is not an exhaustive list*
Specific descriptive statements: • Characterise a numeric percentage or other summary statistic • Identify the sample to which it applies • Identify the outcome of interest	General - Access	In general, students indicated that the [PRODUCT] is accessible and easy to navigate	Any study design capable of supporting correlational, comparative, or causal efficacy statements can also support descriptive statements via analysis of surveys, secondary analysis of administrative data, or cohort analysis with no controls for baseline factors or potential mediating/ moderator variables	Frequencies
	General - Completion	Most students completed the course		
	General - Achievement	Students generally felt the product helped them achieve their goal		
	General - Progression	In general, students felt the product prepared them well for the next level of their learning		
	Specific - Access	X% of learners had a positive experience using the product X% of students are able to access [PRODUCT] on their computer X% of students are able to access [PRODUCT] on their smartphone X% of students could access their assignments easily; and X% of students find [PRODUCT] easy to navigate"		
	Specific - Completion	X% of learners completed the course		
	Specific - Achievement	X% of students indicate that [PRODUCT] helps them to improve their English 'very significantly' or 'significantly' X% of students indicate that [PRODUCT] helps them improve speaking, listening, vocabulary, grammar and writing ***Made up example:*** 86% of test-takers passed the exam in 2015 compared to 50% of test-takers in 2016		
	Specific - Progression	X% of learners employed after completing the credential/course X% of learners progressed to next level after completing a course/module in which Pearson [PRODUCT] or service was used		

Relational - Correlational	Efficacy statement category	Illustrative example statement	Research design used by Pearson associated with statement type	Analysis appropriate for statement *Note: this is not an exhaustive list*
Specific correlational statements: • Include a clear description of the comparison or reference groups, which are defined in terms of the independent variable of interest (e.g. "heavy" Mastering users versus "light" Mastering users) • Specify a dependent variable of interest and identify the metric (e.g. final exam scores) • Specify nature of the relationship (i.e. whether significant or not; whether positive or negative) • Suggest magnitude in terms of an effect size or mean differences along with statistical test statistics • Tech reports document model fit indices (r-squared, pseudo r-squared) where possible	General - Access	Students who spent more time in the etext tended to rate their own engagement in the course more highly	Cohort analysis with controls for baseline factors or potential mediating/ moderator variables	HLM, SEM, Path analysis, other multiple regression analysis techniques, chi-square test, classification, and regression trees
	General - Completion	Learners who made more homework attempts were more likely to complete the course (for binary outcome measures)		
	General - Achievement	Among [PRODUCT] students/users, students who completed more homework assignments tended to earn higher final exam scores than students who completed fewer homework assignments (for continuous outcome measures) Among [PRODUCT] students/users, completion of homework was positively associated with higher final exam score (for continuous outcome measures) Among [PRODUCT] users, learners who accessed more hints were more likely to earn a final grade of B or better (for binary outcome measures)		
	General - Progression	***For binary outcome variables:*** Among [PRODUCT] students/users, time spent on the assignments was positively related to the probability of enrolling in the next level course. Among [PRODUCT] students/users, students who spent more time on the assignments were more likely to enroll in the next level course than students who spent less time on the assignments		
	Specific - Access	***Made up example:*** Students who spent an additional 3 hours in the etext over the course of the semester were 20% more likely to say that they were engaged in the course		
	Specific - Completion	***Made up example:*** Learners completing 30 additional homework problems were 1.5 times more likely to pass the course (for binary outcome measures)		
	Specific - Achievement	After controlling for other factors that may influence achievement, each additional unit of homework completed is associated with/is related to/linked to/connected to an X% increase in students' module/course test score (for continuous outcome measures) ***Made up examples:*** Learners who opened 20 more hints were 2 times more likely to earn a final grade of B or better (for binary outcome measures) Students who scored an additional ten percentage points higher on [PRODUCT] quizzes versus similar students tearned on average 4.33 (±1.10) more percentage points on unit exams		
	Specific - Progression	***For binary outcome variables:*** Among [PRODUCT] students/users, each additional unit of time spent on the assignments was associated with a X% greater probability of enrolling in the next level course ***Made up example:*** Among [PRODUCT] users, students who spent more than 30 hours on the assignments were twice as likely to enroll in the next level course as students who spent fewer than 30 hours		

Comparative (not causal)	Efficacy statement category	Illustrative example statement	Research design used by Pearson associated with statement type	Analysis appropriate for statement *Note: this is not an exhaustive list*
Specific comparative statements: • Specify the treatment group • Specify the comparison group • Use the terms "matched" or "similar" only to indicate variables on which baseline equivalence was achieved. Where baseline equivalence is not reached, those variables should be included in the model as covariates and statements can only reflect that the analyses ""adjusted"" for those variables, not that they were matched or similar. • Indicate whether there was a significant difference between groups. If so, what is the direction of difference? • Suggest magnitude in terms of an effect size or mean differences along with statistical test statistics • Indicate the outcome measure (e.g. final exam scores) • Technical and research reports must include a statement that results cannot be interpreted as causal and identify the specific reason (e.g., failure to achieve baseline equivalence, differential attrition, other confounding factors unable to control for)	General - Access	Learners using the Pearson [PRODUCT] reported a better learning experience than learners using a competitor product / not using the Pearson product - [inc. caveat for statistical adjustment]	QED or RCT, where baseline equivalence was not achieved and/ or where attrition does not meet What Works Clearinghouse guidelines	Analysis of Variance techniques, t-tests, multiple regression, HLM or HGLM
	General - Completion	Learners using the Pearson [PRODUCT] were more likely to complete assignments than learners using a competitor product / not using the Pearson product - [inc. caveat for statistical adjustment]		
	General - Achievement	Learners using the Pearson [PRODUCT] earned significantly higher final exam scores than learners using a competitor product / not using the Pearson product - [inc. caveat for statistical adjustment]		
	General - Progression	Learners using Pearson [PRODUCT] were more likely to progress to college level [SUBJECT] after using the product than those who did not use Pearson [PRODUCT] or used a competitor product - [inc. caveat for statistical adjustment]		
	Specific - Access	Learners using the Pearson [PRODUCT] were X% more likely to report they had a positive learning experience than learners who learned the same math skill without the use of Pearson [PRODUCT] or used a competitor product - [inc. caveat for statistical adjustment]		
	Specific - Completion	Learners using Pearson [PRODUCT] were X% more likely to complete and master assigned tasks/activities compared to learners who do not use Pearson [PRODUCT] or used a competitor product - [inc. caveat for statistical adjustment]		
	Specific - Achievement	There was no statistical difference in percentage scoring proficient in math and reading between student cohorts in Connections Academy schools and cohorts in brick-and-mortar schools that were matched on prior achievement, after adjusting for district-mean student mobility and school-mean student SES and other demographic factors. Learners using the Pearson [PRODUCT] earned X% higher exam scores than learners using a competitor product / not using the Pearson product - [inc. caveat for statistical adjustment] Learners using the Pearson [PRODUCT] showed X% higher course pass rate than learner using a competitor product / not using the Pearson product - [inc. caveat for statistical adjustment]		
	Specific - Progression	Learners using Pearson product were X% more likely to progress to college level [SUBJECT] after using the product than those who did not use Pearson [PRODUCT] or used a competitor product - [inc. caveat for statistical adjustment]		

Causal	Efficacy statement category	Illustrative example statement	Research design used by Pearson associated with statement type	Analysis appropriate for statement *Note: this is not an exhaustive list*
Specific causal statements: • Specify the treatment group • Specify the comparison group • Indicate whether there was a significant difference between groups. If so, what is the direction of difference? • Can only be used where groups were equivalent on achievement at baseline (baseline achievement difference between groups is <.25 SD and if difference is >.05, statistical adjustment is required) • Suggest magnitude in terms of an effect size or mean differences along with statistical test statistics • Indicate the outcome measure (e.g. final exam scores) • For Instrumental Variables or Fuzzy Regression Discontinuity techniques, statements should specifically describe who results apply to, and refrain from generalizing beyond study samples	General - Access	Learners using the Pearson [PRODUCT] reported a better learning experience than similar/matched learners using a competitor product	QED ('Propensity score matching, Regression discontinuity, Instrumental Variables, Fuzzy regression discontinuity) or RCT, where baseline equivalence can be demonstrated and attrition is within What Works Clearinghouse guidelines	Analysis of Variance techniques, t-tests, multiple regression, HLM or HGLM *Note: while these techniques for analyzing group differences may be used in other studies, in all cases in order to make a causal statement, demonstrating equivalence at baseline and random attrition is required*
	General - Completion	Learners using the Pearson [PRODUCT] were more likely to complete assignments than similar/matched learners using a competitor product		
	General - Achievement	Learners using the Pearson [PRODUCT] earned significantly higher final exam scores than similar/matched learners using a competitor product		
	General - Progression	Learners using Pearson [PRODUCT] are more likely to progress to college level [SUBJECT] after using the software than are similar/matched learners who did not use Pearson [PRODUCT] or used a competitor product		
	Specific - Access	Learners using the Pearson [PRODUCT] are X% more likely to report they had a positive learning experience than are similar/matched learners who learned the same math skill without the use of Pearson [PRODUCT] or used a competitor product		
	Specific - Completion	Learners using Pearson [PRODUCT] are X% more likely to complete and master assigned tasks/activities compared to similar/matched learners who do not use Pearson [PRODUCT] or used a competitor product		
	Specific - Achievement	Learners using Pearson [PRODUCT] achieve X% higher on module/course tests compared to similar/matched learners who did not use the software or used competitor courseware		
	Specific - Progression	Learners using Pearson product are X% more likely to progress to college level [SUBJECT] after using the software than are similar/matched learners who did not use Pearson [PRODUCT] or used a competitor product. A significantly higher percentage of students using Pearson product progressed to the next level course compared to students who used competitor product (X% versus Y%).