

设计科学的效度

Validity in Design Science

凯·拉森 (Kai R. Larsen)

科罗拉多大学利兹商学院，美国科罗拉多州博尔德 80309 {Kai.Larsen@colorado.edu}

罗曼·卢基扬年科 (Roman Lukyanenko)

弗吉尼亚大学麦金太尔商学院，美国弗吉尼亚州夏洛茨维尔 22904 {romanl@virginia.edu}

罗兰·M·穆勒 (Roland M. Mueller)

德国柏林经济法学院，商业与经济系 {roland.mueller@hwr-berlin.de}

维达·C·斯托里 (Veda C. Storey)

乔治亚州立大学J. Mack Robinson商学院，美国佐治亚州亚特兰大 30303
{vstorey@gsu.edu}

杰弗里·帕森斯 (Jeffrey Parsons)

加拿大纽芬兰纪念大学工商管理学院，圣约翰，纽芬兰和拉布拉多省 {jeffreyp@mun.ca}

黛布拉·范德米尔 (Debra VanderMeer)

佛罗里达国际大学商学院，美国佛罗里达州迈阿密 33199 {vanderd@fiu.edu}

德克·S·霍沃卡 (Dirk S. Hovorka)

悉尼大学商学院，悉尼大学，澳大利亚悉尼 {dirk.hovorka@sydney.edu.au}

陈淑珍 (Shu-Chen Chen)

弗吉尼亚大学，东亚语言文学文化系，美国弗吉尼亚州，夏洛茨维尔，22904 {sc8u@virginia.edu}

研究人员必须确保其工作产生的知识的主张是有效的。然而，在设计科学中，效度既没有得到很好的理解，也没有得到一致的确立，因为设计科学涉及开发和评估用于解决问题的构件（模型、方法、实例和理论）。因此，证明和传达关于构件的知识主张的效度是一项挑战。本文定义了设计科学中的效度，并推导出设计科学效度框架和应用该框架的过程模型。该框架包括三种高级主张和效度类型——标准、因果和环境——以及效度子类型。该框架指导研究人员将效度考虑因素整合到采用设计科学的项目中，并为日益增多的设计科学方法研究体系做出了贡献。它还提供了一种系统的方法来阐明和验证设计科学项目的知识主张。我们将该框架应用于现有研究的实例，然后用它来证明关于框架本身的知识主张的效度。

Abstract – Researchers must ensure that the claims about the knowledge produced by their work are valid. However, validity is neither well-understood nor consistently established in design science, which involves the development and evaluation of artifacts (models, methods, instantiations, and theories) to solve problems. As a result, it is challenging to demonstrate and communicate the validity of knowledge claims about artifacts. This paper defines validity in design science and derives the Design Science Validity Framework and a process model for applying it. The framework comprises three high-level claim and validity types—criterion, causal, and context—as well as validity subtypes. The framework guides researchers in integrating validity considerations into projects employing design science and contributes to the growing body of research on design science methodology. It also provides a systematic way to articulate and validate the knowledge claims of design science projects. We apply the framework to examples from existing research and then use it to demonstrate the validity of knowledge claims about the framework itself.

关键词：设计科学、设计科学研究 (DSR)、设计科学效度框架、知识主张、研究效度、标准效度、因果效度、环境效度、特征效度、功效效度、外部效度、生态效度、评估、验证

引用 / Cite： Larsen, K. R., Lukyanenko, R., Mueller, R., Storey, V. C., Parsons, J., VanderMeer, D., and Hovorka, D. 2025. 《设计科学的效度 Validity in Design Science》 *MIS Quarterly*, Translated by Shu-Chen Chen, pp. 1-40. English original : bit.ly/4inr9J9

介绍

科学知识必须可信可靠 (Burton-Jones 等人 2021; Creswell 和 Miller 2000)。既定的效度传统有助于科学的可信和可靠，因为它提供了一种系统的方法来评估研究中声称对知识的贡献 (Cook 和 Campbell 1979; Guba 和 Lincoln 1994)。每个科学学科都有模式和程序来评估关于感兴趣现象的知识主张的效度。阐明这些模式可以提供评估清晰度，提高研究效率，促进最佳实践的共享，促进科学的累积，并有助于提高公众对科学的信任。

设计科学是信息系统、计算机科学、工程、医学、管理以及材料和生物医学科学领域的一种研究类型，研究人员对他们所发明的构

件的特性或性能提出知识主张。研究人员已经提出了进行设计科学的流程，每个流程都将评估视为必不可少的步骤 (Gregor 和 Hevner 2013; Hevner 等人 2004; Peffers 等人 2007; Prat 等人 2015; Sein 等人 2011)。然而，现有的方法和框架既没有将知识主张的效度视为构件评估的核心组成部分，也没有就评估有关构件的知识主张的效度意味着什么达成一致。

设计科学展现了知识主张反复出现的模式。示例包括：构件的表现优于现有技术水平 (Padmanabhan 等人, 2022 年)；构件的最新版本比以前的版本性能更好 (Sein 等人, 2011 年)；模型与真实世界系统的参照物相对应 (Gonzalez-Huerta 等人, 2017 年)；构件由于存在特定的设计特征而起作用 (Abb

asi 等人, 2012 年)。设计科学中还有一系列常见的评估方法 (Prat 等人, 2015 年; Venable 等人, 2016 年), 尽管这些方法并未解决将主张与验证联系起来的逻辑, 即确定特定类型的知识主张效度的过程。

虽然某些设计科学项目使用了效度类型, 但这些效度类型通常是从其他传统改编而来的。例如, 内部效度和外部效度改编自心理测量学 (Baskerville 等人, 2015 年; vom Brocke、Hevner 等人, 2020 年), 而效度指标 (如准确率和召回率) 则取自信息检索、统计学和其他学科 (Larsen 和 Becker, 2020 年)。实例化效度 (Lukyanenko 和 Parsons, 2020 年) 是一个例外, 它被提议作为一种本土的设计科学效度类型。

大多数设计科学验证缺乏任何标签, 这使得有效地引用它们并与社区分享最佳实践变得困难。此外, 一些评估程序, 如适用性检查 (Rosemann和Vessey 2008), 回应多个 (通常未明确说明的) 主张, 导致对于究竟验证了什么产生疑问。尽管有潜在的好处, 但尚未有系统性尝试来定义或调查设计科学家所涉及的所有构件类型 (例如, 模型、方法、设计理论; 参见Gregor和Hevner 2013) 的效度, 或提供验证主张的流程。因此, 设计科学中什么构成效度的本质问题仍然未得到解答。

系统性的效度方法将有益于设计科学。首先, 它将促进高质量的学术研究。对于设计科学质量至关重要是知识主张与支持这些主张的证据之间的联系。这可以通过按照公认的结构来遵循评估程序来实现, 这些结构已知可以支持所提出的主张类型。这些程序通常要求程序严谨性和仔细的论证 (例如, 对内部效度的因果关系证据, 通常通过随机实验提供)。确立效度将使主张与其验证之间的联系更加明确, 更容易受到关注和审查。这

种透明度对于产生对主张和报告发现的信任和信心很重要 (Burton-Jones等, 2021)。

其次, 由于特定类型的效度针对特定类型的主张 (例如, 内部效度解决因果关系主张), 为设计科学建立效度传统将使知识主张更加明确。这可以帮助研究者更好地理解 and 交流他们所做的贡献, 并促进验证过程。此外, 阐明的知识主张也可以被其他寻求在设计科学内部和超越设计科学范围扩展原始贡献的研究者使用, 从而使设计科学家能够更好地为关于信息技术构件的设计、使用和影响的累积研究传统做出贡献。

第三, 对效度的全面概念化将增加设计科学对现实世界的影响。通过使研究者对知识主张的性质和效度类型保持敏感, 当构件暴露于意外的现实世界环境中时, 失去其预期目的的可能性会降低。例如, 在机器学习中, 研究者了解哪些构件特征产生了改进的结果是很重要的。然而, 这些仍然狭隘地集中在所提出构件的性能上。一个更具包容性的效度概念将考虑构件对各种现实世界条件的适用性, 包括更广泛的任务和不同的用户群体 (Ethayarajh和Jurafsky 2020)。

第四, 建立效度传统将促进设计科学研究的发表。效度传统是社区规范和共享概念的体现, 它通过增加验证的一致性 (Chan 2014) 和简化验证实践来提高研究生产力。值得注意的是, 高产的计量经济学和心理测量学领域有着独特的效度或验证方法 (Taylor 2013), 定性研究 (Creswell和Miller 2000) 以及计算机科学和软件工程的领域也是如此, 例如模型模拟 (如Pääkkönen等, 2017) 和问题可判定性 (如Fan等, 2018)。跨学科来看, 对核心学科规范 (包括验证) 的更高层次认同与研究生产力和发表质量的提高相关 (Gumpert 2007)。共享学科概念化的正

规化对于加速科学进步至关重要（美国国家科学院 2022）。

最后，确立设计科学中效度的本质将明确其相对于既定传统（例如行为科学）的定位，并更好地向外部人员传达设计科学的独特贡献。验证的共同模式获得系统性的命名，从而帮助设计科学学者和外部人员参考这些评估方法。这反过来又有助于建立设计科学的身份。

本研究做出了几项贡献。我们首先考察了效度概念在设计科学中的使用方式。然后，我们考虑效度的一般本质，以确定可能使设计科学家受益的既定基础。我们通过深入了解主张、构件类型、效度类型和评估背景之间的逻辑和关系来做到这一点。这些基础随后使我们能够定义效度类型，并适当地使用它们来评估设计科学知识主张。

接下来，我们开发设计科学效度框架，以提供一种系统性方法来阐述和验证设计科学项目的知识主张。该框架包括三种高层次的知识主张和效度类型——*标准*、*因果*和*环境*——以及效度子类型。这一结构对应于设计科学的目标，即开发创新构件作为社会挑战的解决方案，同时贡献从业者可以在不同情境中重复使用的科

学知识。该框架在其覆盖的设计科学构件范围上是具包容性的，包括已实施的系统（例如，工具和已部署的实例），封装系统开发蓝图的抽象贡献（例如，概念模型或机器学习方法），以及理论设计知识（例如，规定设计和行动以达到特定目标的设计理论，参见Gregor和Jones 2007）。

最后，我们通过一系列由框架本身递归指导的研究来评估设计科学效度框架。通过两次适用性检查、对效度研究的广泛分析以及对设计科学评估的全面分析，我们展示了我们框架对研究者的实用性、其捕捉现有验证实践的能力以及其简约性。我们以关于设计科学效度的本质和重要性的一般讨论作为结束，推荐如何使用该框架，并确定未来的研究机会。

设计科学方法论的发展

设计科学的方法论的发展 设计科学已经发展了方法论基础，以促进可信和可靠的实用科学知识。这些努力集中在确保构件开发的严谨性以及评估构件和设计知识的实用性和质量。

（具体主题见表1）。虽然构件开发的严谨性已经受到了广泛关注，但评估构件和设计知识的实用性和质量的严谨性却落后了，尽管这是一个主要挑战。

表 1. 设计科学的方法论基础

贡献	参考
系统开发与严谨理论和实证评估的整合	Nunamaker 等人(1991)；March 和 Smith (1995)
评估构件和理论的方法	海夫纳等人。(2004)；维纳布尔等人。(2016)；普拉特等人。(2015)；格雷戈尔和琼斯(2007)；图纳宁等人。(2024)
产生和传播设计科学的方法	佩弗斯等人。(2007)；格雷戈尔和赫夫纳(2013)；约翰内森和佩洪斯(2014年)；Baskerville 等人(2015年)；Avdiji和Winter(2019)；伊瓦里(2015)；格雷戈尔等人。(2020)；图纳宁等人。(2024)
核心理论与构件的一致性	Arazy 等人(2010)；格雷戈尔和琼斯(2007)；库奇勒和瓦什纳维(2012)；沃姆·布罗克，温特等人。(2020)
透明度	沃姆·布罗克等人。(2021)；伯顿-琼斯等人。(2021)；卢基亚年科和帕森斯(2020)；海夫纳等人。(2024)

与实践保持一致	Sein 等人 (2011) ;卢基亚年科和帕森斯(2020) ;海夫纳等人。(2024)
设计理论化	曼德维瓦拉 (2015 年) ;Gregor 和 Jones (2007 年) ;Gregory 和Muntermann (2014 年) ; Lukyanenko 和 Parsons (2020 年) ; Gregor 等人 (2020 年)

建立设计科学严谨性的一个重要步骤是了解现有的评估、它们如何提供互补知识以及它们如何重叠。之前没有研究尝试通过建立可能的效度类型来了解设计科学评估的严谨性。这样做需要了解知识主张、构件和验证方法的类型。

我们回顾了表2中总结的著名设计科学框架期刊文章。表2中的大多数框架主要处理构件开发的严谨性。例如，Sein等人(2011)专注于行动设计研究(ADR)，其中评估发生在真实世界的真实环境中。Venable等人(2016)区分了形成性和总结性评估，着重于评估进行的时间。Gregor和Hevner(2013)考虑了定位和呈现设计科学的方式，以及评估是否发生在开发环境内部或外部。Prat等人(2015)将评估方法组织成具有五个不同评估类型维度的分类法。Baskerville等人(2015)开

发了一个框架，用于突出在整个设计科学项目中可能发生的不同类型的知识生产。Tuunanen等人(2024)提供了一种管理复杂多阶段项目的方法，建议这些阶段("梯队")产生不同的构件，可以在每个阶段进行评估。值得注意的是，效度关注从早期框架中几乎没有考虑(例如，Hevner等人2004;Peffers等人2007)到最近方法论研究中的积极考虑(例如，Baskerville等人2015;Tuunanen等人2024)，随着时间的推移而增长。

表2中的流行框架构成了设计科学中既定的方法论指导。它们为理解开展研究和评估构件的一般方法提供了清晰性，但没有提供关于验证活动的指导，包括如何提供证据来证明知识主张。

表2. 设计科学框架和效度的作用

纸	框架概述	效度和主张	评估
Hevner 等人 (2004 年)	设计科学包括构建和评估构件并将其与环境 and 设计知识库相连接。	最少地强调效度。引用其他研究来说明证据的积累最终将如何确立更广泛的设计科学主张的效度。	提出两个需要评估证据的基本问题：1) 该构件提供了什么效用？2) 该效用如何证明？提出贡献源自通过评估证明的效用。
Peffers 等人 (2007 年)	呈现和评估设计科学的六步迭代过程。	不太强调效度。表明设计科学必须采用“特设论证”(第 50 页)。	研究人员观察并衡量构件对解决方案的支持效果。包括将解决方案的目标与使用构件实际观察到的结果进行比较(例如，通过响应时间、生产的项目、用户满意度调查和模拟)。
Sein 等人 (2011 年)	行动设计研究的四阶段模型。	没有提及效度或知识主张。	主张“真实与同步评估”原则来强调行动设计研究的一个关键特征。
格雷戈尔和赫夫纳 (2013)	设计科学出版的方案，包括评估。	构件评估的标准包括效度、效用、质量和功效。	区分开发环境内部和外部的评估。效用标准与绩效是否转移到开发环境之外有关。
Prat 等人 (2015 年)	评估方法的分类有五个维度：目标、环境、结构、活动和评估。	效度是功效和效果的目标维度的一部分。如果构件正常工作(实现目标)，则获得效度。	重点关注评估中的“什么”(构件系统)和“如何”(使用的方法)。评估涉及构件相对于其他解决方案的优越性的“相对性”。确定典型的评估技术和二次评估。

Baskerville 等人 (2015 年)	探究类型。框架：设计与科学、规律性与个案性。	提供 18 项质量标准。提及内部和外部效度。	提供对探究类型的反思，而不是评估本身。
Venable 等人 (2016 年)	设计科学评估存在于两个维度：人工与自然、形成性与总结性。	效度被视为源自评估的强度以及构件实现目的的程度。	评估的主要目的是评估构件实现预期效用的程度。

设计科学中的一些研究已经认识到建立效度的重要性 (vom Brocke、Winter 等人, 2020 年) , 并提出现有的行为效度类型 (例如内部效度、生态效度) 可能适用于设计科学 (Baskerville 等人, 2015 年) 。在采用基于人工智能的技术 (例如机器学习和自然语言处理) 的设计科学中, 通常会报告精确度、召回率和 F1 分数以及其他混淆矩阵指标 (例如 Abbasi 和 Chen, 2008 年; Li 等人, 2020 年) 。这些不是效度类型, 但与效度相关, 因为它们提供了可量化的度量来确定主张的效度。它们也不适用于所有设计科学构件 (例如设计理论) 。为了建立设计科学的效度, 我们首先广泛考虑科学的效度, 然后将其与设计科学的特定关注点综合起来。¹

研究效度

知识主张应得到验证的基本思想可以追溯到古代 (Carter 2019) 。效度这一术语起源于定量社会科学 (例如, Cronbach 和 Meehl 1955) , 后来得到了更广泛的应用, 因为它被用于不断发展的跨学科信念, 即创建适当的评估 (例如, 评估先前知识, 并商定验收标准) 以确定知识主张的效度。

为了理解设计科学中的效度, 我们回顾了来自信息系统 (包括设计科学) 和其他研究领域 (如计算机科学、社会科学、工程学、医学、法律和人文学科) 关于研究效度的约 7,500 个

文献来源。这项工作产生了 2,418 个候选效度, 构成了迄今为止规模最大的效度综述努力, 其规模超过以往研究一个数量级。当我们验证我们的框架时, 我们会在后文报告这些努力。在此, 我们简要回顾效度在科学中的使用方式, 因为这为我们关于设计科学效度的概念提供了信息。

最初, 效度被狭义地理解为 "我们认为我们正在测量的内容与我们打算测量的内容的接近程度" (Roberts 和 Priest 2006, 第 41 页) 。这种观点将效度集中在测量构件上, 特别是在心理测量学中, 关注的是施用于人类的测试、工具或问卷。随着时间的推移, 定量研究者开发了许多不同种类的效度 (例如, 内部效度、生态效度、区分效度、外部效度) , 这些现在在信息系统行为研究中被广泛使用 (Boudreau 等人 2001) 。使用这些效度已被认为是一种 "专业责任" (Shultz 等人 1998, 第 266 页) 。

定性和解释主义研究人员在其工作中主张独特的效度关注, 强调背景、环境和参与者的重要性, 以及研究人员在创建自然、可信、可确认和可靠的研究过程描述方面的作用 (Lincoln 和 Guba 1985; Onwuegbuzie 和 Leech 2007) 。这一传统中的效度被认为是社会共识的产物, 其中效度基于 "可接受的社区" (Moules 等人 2015, 第 172 页) 。反思这些努力, Creswell 和 Miller (2000

¹附录 A 评估了验证实践的当前状态。

年，第 124 页)指出“普遍的共识”是，效度是研究被接受的基础——这一观点被广泛用于各个学科 (Hoyningen-Huene 2013) ，包括信息系统 (Burton-Jones 等人 2021) 。

计算机科学是一门与信息系统设计科学密切相关的学科，机器学习和人工智能等子领域的研究人员已经采用了一个共同的任务框架，专注于共享基准，以验证相对于最新技术的新产品，如共享排行榜所反映的那样 (Mata damas-Hernández 等人，2012 年) 。该框架催化了计算机视觉和自然语言处理的重大进步。例如，机器学习研究人员已经精心开发了因果分析和消融研究等方法，评估设计好的构件 (例如，Chowdary 和 Kanhangad 2022) 。他们在通用任务框架下采用这些方法来测试知识主张并为共享任务知识做出贡献。然而，尽管取得了令人瞩目的进展，但通用任务框架可能限制了所得模型的实际应用，并不利于评估标准，例如“紧凑性、公平性和能源效率” (Ethayarajh 和 Jurafsky 2020，第 1 页) 。

每种类型的研究，有时甚至是每个领域 (例如机器学习) 都形成了不同的效度传统。正如定性和定量效度关注点不同一样，设计科学效度可以有自己的重点。除了简单地借用心理测量、计算机科学或定性研究中的效度类型之外，还有机会解决设计科学的独特性质，其中研究人员开发构件以实现预期结果并验证有关这些构件的知识主张。然而，其他学科中现有的效度研究对设计科学的效度具有影响，因为它涉及一般效度概念和思想。

我们对效度文献的审查得出了以下发现，这些发现为设计科学效度提供了信息。

首先，研究人员一致认为，系统地评估和验证科学研究是必要的 (Cohen 等人，2013 年) 。即使那些认为“效度”标签不足以证明其研究成果的研究人员也承认，提炼和分享成功且行之有效的验证模式大有裨益 (例如，Creswell 和 Miller，2000 年；Maxwell 1，1992 年) 。

其次，科学寻求推进构成研究贡献的知识主张 (Collier-Reed 和 Ingerman 2013) 。一项研究可以有許多知识主张，其中一些是研究的重点 (主要主张)，另一些则来自先前的研究 (次要主张)。通常，只有前者需要在研究中进行验证。因此，在效度背景下，**知识主张**是对感兴趣现象的断言，它捕捉了研究的原始贡献。

第三，在验证一项知识主张时，通常将主张的对象与一个**参考实体** (抽象或具体的对象) 进行比较，其属性或行为可与所提出的想法或对象进行比较，以评估后者的质量。例如，在计算机科学中，当“模型将真实系统的输入输出行为模仿到某种可接受的准确度水平”时 (Murray-Smith，2015，第 30 页) ，以真实系统为参考实体，即可实现功能效度。参考实体以多种形式出现，包括构件、现有的自然对象 (例如人类) 和心理观念。无论研究主张如何，总是存在一个参考实体可以用来验证该主张。即使是对于激进的创新或发明，也有参考实体可以与之进行比较。²

第四，为了验证知识主张，研究人员进行一项或多项评估 适合该主张的程序。 评估

² 想想人类历史上的伟大发明，包括火、轮子、炸药、钉子和印刷机。所有这些发明都有手工或创新性较低的构件，可以作为评估的依据。虽然许多革命性发明的天才之处在于它们与现有制品或执行任务的方式 (参考实

体) 的偏离程度，但当无法识别现有参考制品或无法访问 (例如，它是专有的) 时，可以考虑其他相关制品或流程。

(验证)程序是为提供知识主张效度的证据而采取的一系列任务。通常,该程序涉及在某些维度上将焦点实体(例如,设计科学研究中开发的构件)与参考实体(可能是物质的或抽象的)进行比较。这些程序通常是由学科内部的共识建立的(泰勒 2013)。

第五,效度是一个程度问题,因为参考实体的适当性和质量以及验证它的方法可能会有所不同。因此,特别强的比较是保守的——根据最佳知识、最先进的构件或做事方式(称为标准)进行。定量和定性研究人员(例如,Moules 等人,2015年;Taylor,2013年)都认为,实现完美的效度是不可能的。每个研究团体都会确定构成比较充分结果的标准,以便相应的知识主张可以被接受为科学有效。

最后,为了系统化商定的和广泛使用的参考实体模式、评估程序和接受特定类型知识主张的证据的规范,对它们进行标记和组织成**效度类型**。然后,主张某种特定的效度(例如,收敛效度)可以作为一种捷径,表明相应的主张(即两个测量值实际上是相关的)可以根据现存的科学规范被接受为有效的。例如,库克和坎贝尔(1979)研究了威胁并提出了确保社会科学内部效度的方法,而林肯和古巴(1985)提出了证明定性研究中主张的可信度和真实性的步骤。这些效度模板提出了识别参考实体、比较它们和呈现研究结果的具体方法。这使得引用和共享验证实践、一致使用它们以及通过识别效度类型之间的联系、重叠和差距来改进它们变得更加容易,从而有助于形成一种累积的传统。

设计科学效度框架

考虑到效度的基础,我们现在开发了设计科学效度框架。³我们从关于焦点构件的知识主张开始。阐明这些主张可以帮助学者共享和重复使用有用的验证程序。

设计科学知识主张

设计科学主张的本质根植于设计科学作为功利性科学探究(Hevner 等人,2004年)。设计科学的一个关键目标是创造解决现实世界挑战的构件,并产生与这些构件相关的设计知识。例如,Hevner 等人(2004年,第77页)将设计科学定义为“创造和评估旨在解决已确定的组织问题的IT构件。”从那时起,对设计科学的理解已经超越了纯粹的组织重点,因为研究人员越来越多地寻求解决更广泛的社会和个人挑战(Weinhardt 等人,2020年;Winter 等人,2014年)。此外,许多构件是社会技术的,将软件、硬件和流程与个人、团体和组织结合在一起(Thomas 等人,2022年)。因此,设计科学涵盖了广泛的社会问题和相应的创新构件,这是我们采用的包容性视角。设计科学与实践的区别在于开发与建造构件相关的知识的目标(Gregor 和 Jones,2007年;Peppers 等人,2018年)。因此,我们将**设计科学定义为**开发新颖的构件和相关设计知识来应对个人、组织和社会挑战和机会的研究。

设计科学中有许多类型的成果。常见的成果有模型、方法、实例、设计理论和理论组件(例如构造、设计原则)(Gregor and Hevner 2013)。成果是具有多个相互关联部分或组件的复杂实体。⁴它们通常也被视为更大

³ 在下文中,我们将设计科学效度框架简称为“效度框架”,除非我们有意使用其全名来提及它。

⁴ 每种构件类型又有许多子类型。例如,模型子类型包括框架、分类法、本体论、模拟和数学模型。

系统的组件，例如组织或更广泛的技术系统。考虑到这种多样性，我们将**设计科学成果定义**为作为设计科学项目的重点贡献而创建的抽象或具体实体，旨在实现个人、组织或社会层面的预期成果。

设计科学为问题提供解决方案并产生规范性设计知识的基本原则是设计科学知识主张的核心。具体而言，要评估给定构件的贡献，该构件必须在某些方面优于现有解决方案（例如更快、更高效、更便宜）（Padmanabhan 等人，2022 年），并且必须指定用于创建构件的方法（Hevner 等人，2004 年）。要发挥科学的作用并让其他人从积累的设计知识中受益，研究人员需要传达有关构件如何构建的知识——例如，通过在存储库中共享代码或在概念层面描述构件如何运行（Burton-Jones 等人，2021 年；Hevner 等人，2024 年）。

此外，可以提供将设计选择（构件的各个部分）与期望结果关联起来的具体因果机制，以深化从构件开发中产生的设计知识（Gregor 和 Jones 2007；Peffers 等人 2018）。最后，由于从业者不太可能在与文章中提出的原始研究环境相同的环境中实施他们的解决方案，因此指定构件预计何时、在什么条件和界限下实现其结果是很有价值的（Hevner 等人 2024）。设计科学家的一项关键任务是支持其他研究人员和从业者在新环境中重复使用构件和设计知识（例如，Iivari 等人 2021）。

开展设计科学活动会产生特定类型的知识主张。**设计科学 知识主张**是关于一件构件的命

题，它通过特定的形式或功能宣称其对科学和社会的贡献。例如，Tiefenbeck 等人（2016）开发了一种淋浴计量表，并声称与缺乏此类指标的淋浴相比，水和能耗的特定指示（构件的形式）降低了能耗（构件的功能）。⁵与设计科学的目标一致，设计科学知识主张提出，通过特定方法构建的构件会带来预期的结果，结果是由构件的特定设计属性引起的，或者结果预计会在特定环境中或不同环境中保持不变。我们分别将这些类型的知识主张标记为**标准**、**因果**和**环境**，并在下文中进行详细说明。

标准主张：在我们审查的每篇设计科学论文中，作者都会对构件的**预期结果提出一个或多个主张，例如其在应对挑战或机会方面的好处或效用。我们将标准主张定义为关于构件效用的知识主张。做某事总是存在现存的或替代的方法，因此标准主张隐式或显式地呈现与此类现有实体或流程（也称为最先进技术或标准）的比较**（Padmanabhan 等人，2022 年）。例如，作者可能声称开发了一种更有效的构造搜索引擎（Li 等人，2020 年）或一种可定制技术的新颖设计理论（Germonprez 等人，2007 年）。这些主张分别陈述或假设，传统搜索引擎可以通过从论文中提取理论构造来更好地工作，或者典型的设计理论不涉及可定制技术。我们将这些主张称为**标准主张**，即使主张本身不需要明确指定标准。

当比较对象是公认的最先进技术时，标准主张尤其有力（Padmanabhan 等人，2022 年）。例如，Larsen 和 Bong（2016）将他们的搜索

⁵ 在设计理论的背景下，Gregor 和 Jones（2006 年，第 327 页）提出了“可测试命题或假设”的相关概念，解释说这些关于设计理论的命题是“通过实例化、构建系统或实施方法，或者在极少数情况下通过演绎逻辑”进行测试的。设计知识声明是一个更广泛的概念：正如

我们在论文中所展示的，知识主张适用于任何构件，并且为了得到验证，它们不需要实例化其组件，即使在设计理论的情况下也是如此。

算法的性能与 EBSCO Host 搜索引擎标准进行了比较，而 Li 等人(2020)通过将他们的搜索算法与 EBSCO Host 和 Google Scholar（公认是学术搜索的最先进技术）进行比较，提出了一个更有力的主张。在我们对设计科学文章的检查中（稍后报告），我们发现所有论文都有标准主张，隐含或明确地表明结果比现有构件或流程更好。

因果主张：从长远来看，标准主张不足以确保科学进步。在标准主张确立后，研究人员可能（在确立标准主张的研究中或在后续研究中）希望通过证实关于构件的哪些设计特征产生其声称的结果的主张来深化设计知识。我们将此类主张称为*因果主张*，鉴于科学对因果关系的基本兴趣（Salmon 1998）。因果主张是一种知识主张，关于构件的特定部分在多大程度上导致特定效用。

虽然提供构件的代码或描述有助于支持构件的价值并使其更易于复制和扩展，但提出因果主张可以帮助构建更好的构件并开发新的技术理论。例如，Abbasi 和 Chen (2008)在指出CyberGate 的功能集时提出了几个因果主张，比基线特征集更能表示信息类型，在之前的系统中很常用（请注意，也存在一个标准主张）。他们对每个扩展特征集（即主题、观点、风格、类型和交互信息）进行了单独的评估，以确定每个特征集对构件性能的因果影响（标准来自采用所有特征的构件的性能）。

因果论断并不局限于机器学习，甚至不局限于实例化的构件。在我们的评估中，我们评估关于我们自己的设计科学效度框架各部分价值的因果论断。事实上，鉴于大多数类型的设计科学构件都不是实例化的，定性方法在评估这些构件方面发挥着重要作用。反过来，这些方法可以基于论证知识（例如，核心理论），这些论证知识表明了将构件特征

与目标结果联系起来的因果机制（Gregor and Jones 2007）。

我们并不总是能够知道某件事物是如何或为何运作的，尤其是在处理复杂技术、创新构件和“发明”时(Gregor and Hevner 2013)。因此，并非所有论文都提出了这样的主张；在我们对设计科学论文的调查中，50% 的论文提出了因果主张。

环境主张：构件总是在某种环境下创建，以应对给定的挑战或机会。然而，环境的作用在设计科学中并不总是被考虑。为了支持从业者实施构件，*环境主张*阐明了构件预期结果预期在哪些情况或条件下成立。因此，当在多个环境或更像构件预期使用环境的环境下进行评估时，标准和因果主张通常被认为更为严谨。

并非所有设计科学论文都会提出环境主张。行动设计研究是一个例外，其中实施构件的环境，即真实的组织(Sein 等人，2011)，是任何有关构件的主张旨在适用的环境。可以通过暗示构件在其原始环境之外发挥作用（例如在相似甚至不同的环境中），来扩大环境主张（从而加强标准或因果主张的范围）。此外，环境主张有助于描述研究结果的概括。例如，Lukyanenko 等人（2019）明确声称他们的基于实例的数据收集方法对于大型公民科学项目中的开放式任务特别有效，但也声称它预计不会为微任务众包中的封闭式收集带来好处。

在我们的设计科学论文样本中，14% 的论文提出了环境主张。这个数字可能被没有明确提出此类主张的作者压低了，即使这些主张很可能得到提供的证据的支持。如果是这样，这对研究人员来说是一个考虑推进环境主张的机会（以便其他阅读他们作品的人能够更

好地理解他们开发的解决方案在其他环境下也可能有用)。

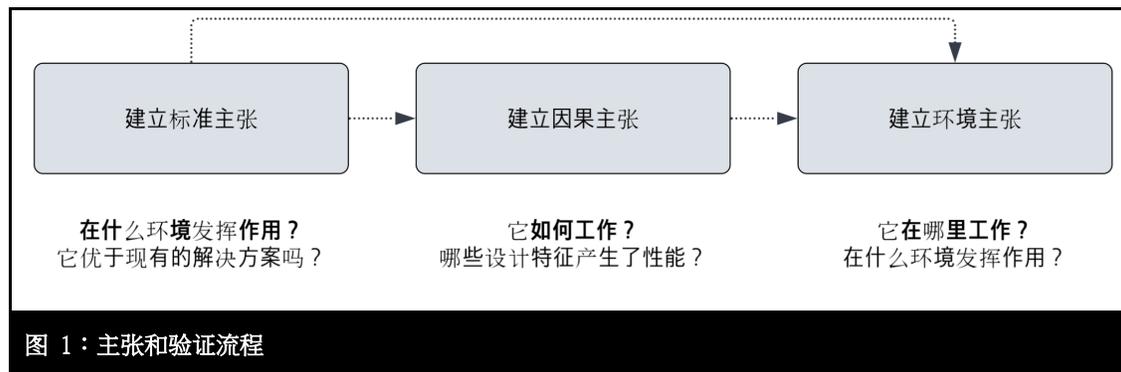


图 1 总结了基于知识主张的重要性的验证过程。该图表明每篇论文至少提出了一个标准主张。这对于新颖的构件来说可能就足够了。正如 Gregor 和 Hevner (2013, p. 346) 所建议的那样，新颖的构件是“有趣的应用，目前对问题环境的理解很少，并且没有有效的构件可用作解决方案。”然后，根据构件的新颖性，论文可能会额外提出因果和/或环境主张，以加深对构件为何和在何处起作用的理解。因果主张需要同时或先前建立标准主张；环境主张需要建立标准主张或因果主张。

三种类型的知识主张可以针对所有类型的构件做出。如前所述，抽象构件（例如，设计理论、概念模型）可以实例化，并且（通过实例化）可以显示对世界具有可衡量的影响。此外，可以使用定性技术（例如反事实分析和利益相关者访谈）来评估抽象实体的标准、因果和环境主张。主张的类型由所需的知识贡献而不是构件类型驱动。

设计科学构件和比较实体

我们定义**设计科学效度**是指关于焦点构件的知识主张得到证据支持的程度。与这三种知识主张类型一致，设计科学效度也分为三种

一般类型：标准效度、因果效度和环境效度。通过支持各自的主张，这些效度类型在三个重要方面推动了设计科学的实践。首先，**标准效度**用于支持关于设计和使用构件的结果的主张，旨在带来期望的改变。其次，**因果效度**用于支持关于特定设计（构件特征、部件）对构件效用的贡献的主张。第三，**环境效度**涉及知识主张对预期环境或其他环境的成立程度。在这些大类中，我们提出了更具体的效度类型，主要基于参考实体的类型，其次基于焦点构件和参考实体之间的比较性质。

这设计科学效度的重点在于对焦点构件的主张。有些文物是**物质的**，有些是**抽象的**（Gregor and Hevner 2013, 第 341 页）。例如，Tiefenbeck 等人（2016 年）的淋浴计量器是物质的，而理论则是概念和命题的系统（Gregor and Jones 2007），是一种抽象的构件。

抽象构件可用于开发诸如淋浴计量器之类的物质构件。通常，有两个**独立但相关的构件**——理论（例如设计命题）和由此产生的物质构件（例如淋浴计量器）。构件是更大系统的组成部分，例如人与构件的**社会技术系统**（Chatterjee 等人，2021 年；Winter 等

人，2014 年）。对于许多项目而言，考虑构件所嵌入的更广泛的系统非常重要，因为知识主张可能是针对系统影响提出的。如果研究人员试图将社会技术系统本身设计为构件（参见 Thomas 等人，2022 年），则可以对这个社会技术系统提出标准、因果和环境主张。⁶同时，如果重点关注的是嵌入社会技术系统（例如，智能城市）中的构件（例如，智能淋浴计量器）的属性和行为，则系统会为构件提供实施设置，从而可以对构件进行环境声明。请注意，焦点构件和比较参考实体存在于（通常不同的）社会技术系统中，从而产生环境效度挑战。

为了推导出参考实体的类别，我们考虑了设计科学产生的主要构件类型——构造、模型、方法、实例和设计理论，以及可用作标准的现有自然物体（vom Brocke、Winter 等人，2020 年；Gregor 和 Hevner，2013 年；Hevner 等人，2004 年；March 和 Smith，1995 年）。表 3 概述了可与焦点构件进行比较的参考实体列表。有了参考实体，我们就可以阐明效度类型，并形式化研究人员可以用来支持其知识主张的特定验证程序。

表 3. 参考实体的类别

类别	子类别	细节
实例化实体 产出实体。使用时，它们会改变世界，这些改变可以被测量和记录。参考实体的产出可能在过去、现在或将来。除产出之外，还可以与其他方面比较。	标准实例 现实世界的实体成为与焦点构件进行比较的标准。	类别包括标准构件（例如，搜索引擎 Google Scholar）、现实世界对象（例如，人类国际象棋大师）、现实世界流程（例如，雇用员工）或报告现实世界状态的传感器（例如，温度或降雨量）。 研究人员有责任证明标准实例可作为比较的合理标准。 实例化可以将构造、模型、方法和设计理论操作化为具体解决方案（例如应用程序、平台、企业系统）。 标准实例用于证明焦点构件的优越性能，或焦点构件近似标准实体的输出、结构或特征的能力。
	操纵构件 从焦点构件中衍生的构件；通常在同一研究中开发，并经过操纵以便通过比较两个构件得出结论。	通过移除或替换焦点构件的一部分来构建。
未实例化的实体 比较的对象是具体物质构件的蓝图。	理论 （及其组成部分：概念和设计原则） 旨在解释、预测或指导行动的概念体系。	设计理论及其组成部分（例如概念、设计原则）和非设计理论（例如解释和预测理论），可以为焦点构件提供设计相关知识。

⁶ 一旦将客户关系管理（CRM）系统（焦点构件）置于组织更广泛的社会和物理系统中，我们就可以评估包括组织在内的更大系统的属性（例如，客户满意度、公平性、销售额、利润）。此外，

通过比较采用 CRM 的组织与未采用 CRM 的组织或引入 CRM 前后的同一组织，我们可以验证有关 CRM 或其部件对感兴趣结果的影响的说法。

除非实例化，否则这些实体不会产生实质性输出。它们通常需要人类专家等代理的解释。	模型 为了增进理解或促进行动而创建的某些现实方面的概念性表征。	模型表征现实的某些方面；通常会忽略与建模者无关的方面；可能会引入有目的的偏见。 模型表征发生设计干预的领域，或者存在其他设计构件，并根据实用性或与事实的对应性进行评估。
	方法 用于完成某项任务的一套独立的、合乎逻辑的步骤序列。	方法可以存在于各种不同的领域。例如，需求分析、流程控制和检查表。

效度类型

标准效度类型：如图 1 所示，所有设计科学项目都应提出标准主张，以表明所设计的构件提供了一些好处。这样做需要采用标准效度类型。标准主张（以及因果主张，如我们稍后所示）可以通过两种方式进行比较：通过其功效或通过其特征。功效比较考虑构件的输出与其他输出生成实体的相似性，而特征比较评估构件及其参考实体的特征之间的相似性（表 3）。因此，标准效度有两种子类型：标准功效效度和标准特征效度。

焦点构件和参考实体具有可比的输出时，标准效度可支持关于实例化构件（表 3）的主张。因此，标准效度处理的标准主张是通过比较焦点构件的效度与被认为代表标准的实例化实体的效度来支持的。这些类型的效度支持以下主张：构件的输出（或此类输出对社会技术系统的影响）相对于参考实体的输出或影响具有效用。

根据时间是否重要，可以区分两种标准效度子类型。如果重要，预测效度（在参考实体生成结果之前对预测进行适当注册）将产生最强的声明效度。然而，在设计科学中，预测效度通常是根据构件创建时存在但直到验证后才提供给构件的未来数据进行评估的。如果时间不是预测的一个因素，未来数据不可用，或者较弱的验证就足够了，则可以评估同时效度。当没有提出预测声明时，可以

采用同时效度来检查焦点构件相对于参考输出的输出。通常，可用案例通过交叉验证分为“训练”、“验证”和“测试”集，其中焦点构件相对于“验证”和“测试”集中的真实值的表现是同时效度的类型。在非正式语境中，术语“标准效度”可能被认为是指同时效度。

标准特征效度处理标准主张，即将焦点构件的特征与被认为代表标准的参考实体的特征进行比较。对于特征效度，代理（通常是人类）通常参与评估，因为：参考构件不产生输出（如理论、模型和方法的情况）；输出不能直接与参考实体进行比较（如比较两个生成式人工智能的输出时的情况）；构件是实例化的，但构件的特征（例如其界面）包含贡献；或者构件与受访者对相关构件或过程的体验进行比较（例如对感知有用性的评估）。所有这些情况都是采用标准特征效度的情况。

特征效度允许评估功效评估无法评估的构件方面。它们更灵活，能够比较不相同的输出或无法自动比较的输出（例如，对医学本体或网站的有用性的看法）。无论评估是基于对输出的直接比较还是构件的特征，特定的效度类型都基于比较实体（表 3）。例如，标准实例效度（实例效度）效用效度（简称“效用”）是指相对于实例化参考实体的构件主张的支持程度。Abbasi 等人（2018 年）使用标准实例效度来验证效用主张。作者要

求组织估计实施该系统与继续当前业务流程相比所实现的节省。

类似地，理论效度涉及关于焦点构件相对于理论构件的主张得到多大程度的支持。例如，Lukyanenko 等人(2019)开发了一个公民科学平台，其设计特征他们声称与设计理论中的相关原则相对应。

标准模型效度（模型效度）是指焦点构件与模型的一致程度。虽然这种效度类型并未出现在我们审查的设计科学论文样本中，但我们在文献综述中考虑了标准模型效度。例如，Refsgaard 等人（2006 年，第 1596 页）建议，在开发模拟模型（焦点构件）的代码时，必须确定“模型代码在某种意义上是现实世界系统（例如生态系统）概念模型（例如专家开发的模型）的真实表示”。

最后，标准方法效度（方法效度）将一种方法与现有方法进行比较，作为用于完成任务的独立、合乎逻辑的步骤序列。标准方法能够根据其他实体（自然或构件及其部件）评估焦点构件或焦点构件的一部分。Piramuthu 和 Doss（2017）在评估他们的构件时提供了一个例子——一种用于同时认证多个射频识别标签的协议。他们使用形式化证明来验证该协议是否满足最高的安全要求（Avoine 等人，2009 年）。

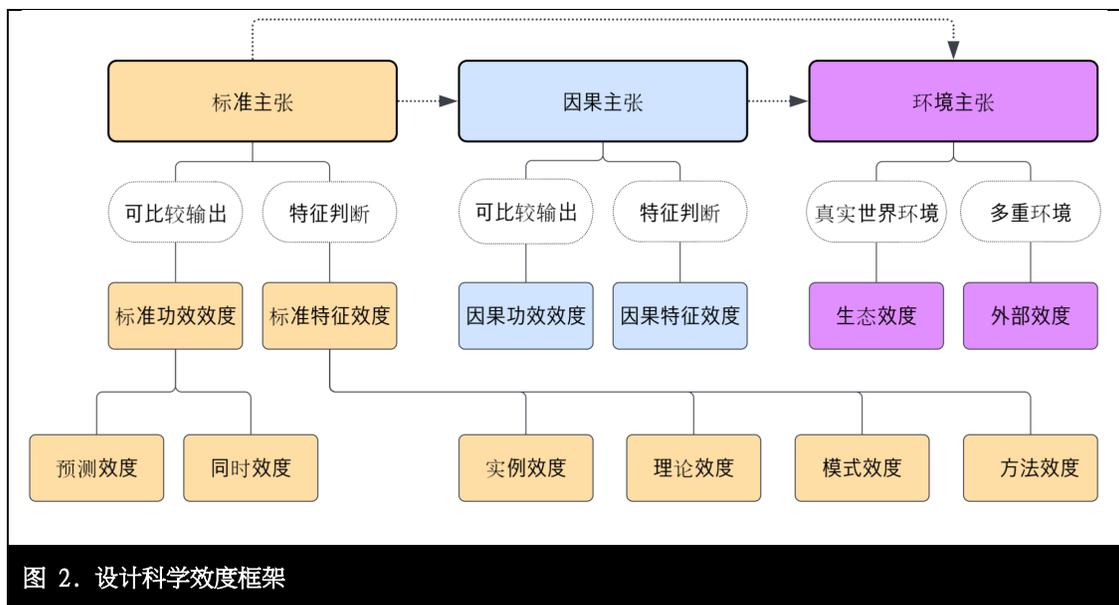
因果效度类型：与标准效度一样，因果效度也有两种子类型：因果效度和因果特征效度。因果效度处理通过评估支持的因果主张，该评估与具有故意不同部分的构件操纵版本的效度有关（与焦点构件相比）。操纵的构件通常由同一研究团队在同一研究环境下开发，并被操纵以允许从两个构件的比较中得出推论。这有时被称为消融研究，由 Newell（1975）引入，并且常见于机器学习中（例如，Ab

basi 等人 2012；Abbasi 和 Chen 2008；Etudo 等人 2017）。

因果特征效度处理的是通过与具有故意不同的组件的操纵构件的特征进行比较来支持的因果主张。实现这种差异的一种方法是删除或更改组件（部分）。与因果效度一样，比较实体通常由同一个研究团队在同一研究环境下开发，并经过刻意操纵以便通过比较两个构件得出推论。例如，Umapathy 等人（2008）开发了两个版本的焦点构件：一个允许在言语行为支持下选择整合模式（焦点构件），另一个则不支持（参考构件）。研究人员进行了一项实验，其中为参与者提供业务流程设计任务（带有特定任务元素列表），以生成企业集成的修改后的业务流程模型和符号（BPMN）模型。专家对模型进行了评估，并比较了感知到的错误数量。

另一种方法依赖于反事实推理（Collins 等人，2004 年）。将焦点构件与移除某个组件的版本进行比较，可以验证该组件不会对构件的性能产生因果影响，并且可以出于简约目的将其移除。

环境效度类型：环境效度类型支持关于评估构件的环境和条件（例如，实验室与目标用户或其代理人使用的现实世界社会技术系统）的主张。这里的目标是评估与构件相关标准和因果主张在某种程度上类似于目标社会技术环境或在不同环境中具有概括性的环境中得到支持的程度。因此，环境效度评估程序要处理条件和环境的差异。环境主张可以针对构件评估的环境（相对于目标社会技术系统）或其他环境。这导致了效度的两种子类型：生态（关注原始评估环境相对于目标社会技术系统的属性）和外部（关注在其他社会技术系统中的评估）。



评价社会技术系统与构件预期使用的社会技术系统之间的相似性越高，主张的生态效率就越高。一些设计科学方法，如行动设计研究，往往具有较高的生态效率，因为评估是在构件使用的自然组织环境中进行的。例如，Zaitsev 和Mankinen（2022）与柬埔寨的参与者合作开发了一款可在现实环境中使用的应用程序。

外部效率被定义为标准或因果知识主张在多于一个社会技术系统中得到评估支持的程度。当评估焦点构件的社会技术系统不同且焦点构件保留其效用时，可以实现更高的外部效率。例如，Zaitsev 和Mankinen（2022）使用行动设计研究方法开发了一款应用程序，

通过培训提高金融知识水平。最初的开发发生在柬埔寨，但该作品后来在尼泊尔进行了改编和使用，提供了外部效率，作者得出结论：“最初的设计已经灵活而简约，根据设计原则创建，为本地化提供了良好的基础”（第 106 页）。

设计科学效度框架及其应用

图 2 显示了设计科学效度框架。该框架通过特定类型的比较，建立了关于焦点构件的三种类型的主张与用于提供支持这些主张类型证据的特定效率类型之间的关系。

表 4 正式定义了每种效率类型。

设计科学效度	关于焦点构件的知识主张得到证据支持的程度。
标准效度	与被认为代表标准的参考实体相比，通过评估支持焦点构件的知识主张的程度。
标准功效效率	与被认为代表标准的物质参考实体的功效相比，关于焦点构件功效的知识主张在多大程度上得到评估的支持。
预测效率	关于焦点构件功效的知识主张通过评估其准确性与在用于创建构件的数据之后产生的参考功效比较得到支持的程度。

同时效度	通过与用于创建构件的数据在同一时期出现的参考输出相比对焦点构件的功效进行评估，可以确定关于焦点构件功效的知识主张在多大程度上得到支持。
标准特征效度	通过对焦点构件的特征与被认为代表标准的参考实体的抽象特征进行比较的评估，可以得出关于焦点构件的特征的知识主张得到多大程度的支持。
实例效度	通过对焦点构件的特征进行评估，将其与被认为代表标准的实例化参考实体的抽象特征进行比较，可以确定关于焦点构件特征的知识主张在多大程度上得到支持。
理论效度	通过对焦点构件的特征与被认为代表标准的理论参考实体的特征进行比较的评估，可以得出关于焦点构件的特征的知识主张得到多大程度的支持。
模型效度	通过对焦点构件的特征与被认为代表标准的模型参考实体的特征进行比较的评估，可以得出关于焦点构件特征的知识主张得到多大程度的支持。
方法效度	通过对焦点构件的特征与被认为代表标准的方法参考实体的特征进行比较的评估，可以得出关于焦点构件特征的知识主张得到多大程度的支持。
因果效度	与缺少该部分的操纵构件相比，关于焦点构件的一部分的影响的知识主张在多大程度上得到评估的支持。
因果功效效度	关于焦点构件某一部分对该构件功效影响的知识主张通过评估其功效与缺少该部分的操纵构件的功效比较得到支持的程度。
因果特征效度	通过对焦点构件某部分的特征与缺少该部分的被操纵构件的特征进行比较的评估，可以得到关于焦点构件某部分特征的知识主张的支持程度。
环境效度	在特定环境中通过评估得到支持的程度。
外部效度	标准或因果知识主张在多大程度上通过多个社会技术系统的评估得到支持。
生态效度	标准或因果知识主张在多大程度上通过与构件所针对的现实世界社会技术系统相对应的社会技术系统中的评估得到支持。

一般来说，不同的效度类型是合适的，这取决于研究人员希望提出的主张类型。正如我们的框架所示，有三种主要的主张类型，与三种主要的效度类型相对应。因此，我们的框架允许研究人员根据关于构件的知识主张的性质来确定适当的效度类型。一旦研究人员决定了一种主张类型，他们就应该从顶部开始遵循该主张类型，直到找到合适的底层效度类型来支持该主张。

例如，Koornneef等人(2020)提出了一种方法来改进识别相关信息和解决航班间飞机故障排除/维护问题的选项的过程。作者声称他们的原型支持更快地检索与问题解决相关的信息（标准声明，见图 2）。他们将原型与行业中当前在维护手册中搜索相关信息的做法进行了比较。由于原型和手动搜索方法提

供了可比的输出（即相关问题信息），因此这是具有特定类型的同时效度的标准效度的一个例子（因为比较时间被认为不起作用）。为了进一步支持他们的说法，作者进行了一项实验，其中要求高级维护培训师使用原型或手动方法查找相关信息。作者测量了两种情况下的搜索时间，发现原型提供了更快的搜索时间。如果评估支持该声明，则可能只需要一种叶节点效度类型。如果作者认为他们的成果不够新颖，不足以保证在目标媒体上发表，他们可以通过提出因果关系主张来加深贡献。虽然Koornneef等人(2020)已经进行了功效效度评估，但在假设的情况下，他们可能会通过评估其成果的哪些部分对检索相关信息的速度贡献最大来增加因果功效效度。

表 5. 示例验证

构件和构件类型	知识主张	效度类型	评估
---------	------	------	----

Meth 等人 (2015) 提出了一种需求挖掘工具，一个实例。	使用该工具可以让用户比不使用该方法时更准确地识别需求。	同时效度	进行实验来评估用户是否使用该工具准确地识别所有需求。
Ramakrishnan 等人 (2023) 提出了平台驱动的知识共享的设计原则，一种理论。	原则具有可及性、重要性、新颖性和洞察力、可操作性和指导性、效度。	理论效度	焦点小组评估设计原则。
Sedrakyan 等人 (2017) 提出了一种包含反馈的快速原型制作 (FIRP) 方法。	FIRP 模拟提高了对模型行为方面的理解。	模型效度*	析因实验设计评估方法的可理解性。
Valecha 等人 (2013) 提出了一种用于应急响应领域消息标准化的共享词汇表模型。	模型完整且正确，可以满足紧急响应人员的通信需求。	模型效度	与应急响应专家进行焦点小组讨论。
*注意：参考实体可能与构件类型不同，因为它涉及建立基线。例如，参与者可以利用他们的世界模型来评估所提出方法的某些方面。			

这个例子说明了该框架如何代表现有的设计科学实践，以及如何回顾性地使用它来分析现有的验证并更好地理解为什么做出某些选择。它还可以提出替代方案（例如，通过要求人类判断来评估原型，从而产生特征效度类型），因此展示了使用该框架指导设计科学验证过程的可能性。

框架的评估和应用

为了评估该框架并同时演示如何使用它，我们将该框架递归地应用于自身。框架（一种模型）应该具有实用性，这意味着，例如，它必须被潜在用户视为有用（Hevner 等人，2004 年；March 和 Smith，1995 年）。框架还应该简洁，同时忠实地捕捉关键领域的概念（美国国家科学院，2022 年）。

我们制定了关于效度框架的几项知识主张，然后对其进行了验证。我们提出了几项标准主张，然后是环境和因果主张，以说明所有类型的主张和不同类型的效度的用法。首先，我们主张框架效用，正如其潜在用户——设计科学家所认为的那样。这些是标准和环境

我们通过考虑设计科学文献中的样本工作来进一步展示该框架的范围。我们通过考虑 March 和 Smith (1995) 以及 Gregor 和 Hevner (2013) 提出的构件类型来确定此演示的范围：实例、设计理论、方法和模型。对于每种工件类型，表 5 都提供了一篇示例论文，并确定了作者提出的构件、主张和效度。

主张，由模型效度和生态效度的证据支持。其次，我们主张框架在系统化现有验证程序（命名和未命名）方面的效用，由模型效度的证据支持。第三，我们主张框架在系统地捕获和组织跨学科提出和定义的现有设计科学效度类型方面的效用。这些是标准和环境主张，由模型效度和外部效度的证据支持。最后，我们主张该框架的组成部分足以实现效用，由因果特征效度的证据支持。

主张 1：标准主张和目标用户的环境主张

对于主张 1，我们通过以下方式评估标准和环境主张：两种适用性检验（Rosemann 和 Vessey 2008），一种常见的设计科学评估

类型（例如 Li 等人 2020 年；Lukyanenko 等人 2019 年）。适用性检验是一种涉及多种效度类型的综合评估。但是，对于评估效度框架，适用性检验对于解决模型效度是有效的，其中参与者将我们的框架与他们的理想框架模型和其他相关框架进行比较。由于适用性检验会吸引框架的潜在用户并要求他们在工作中使用它，因此它还提供了生态效度的证据。我们的评估是在具有丰富评估经验和接触现有评估框架的设计科学家的环境下进行的。

先前的评估框架，包括 Prat 等人(2015)和 Venable 等人(2016)的框架，分别侧重于评估模式以及同时和事后评估策略。通过对评估技术和标准的六个维度进行编码，Prat 等人。确定用于评估的具体技术（例如，观察、描述、实验、动态分析）和具体研究方法（例如，定性、实验、模拟、指标）。Venable 等人（2016 年）的框架侧重于进行评估的人工/自然环境，以及这些评估如何具有形成性（在项目期间进行）或总结性（在项目结束时进行）。

相比之下，我们的效度框架为通过评估过程验证知识主张提供了潜在的推理。我们指的是通过评估过程获得的证据在多大程度上支持研究人员对焦点构件所作的知识主张。此外，我们的分析表明，设计科学中的评估总是比较：将绩效标准与现有标准构件（或实现相同结果的现有方法）进行比较，将期望的构件特征与现有构件进行比较，或在构件应用环境内和/或环境之间进行比较。在阐明标准、因果和环境效度之间的关系时，我们的效度框架超越了先前的评估框架。我们展示了执行评估的过程，并说明了如何建立知识主张类型的效度。通过明确这些关系比较，我们的框架：为评估提供元类别（例如，标准、因果和背景）；将现有的效度组织到

这些元类别中；并确定评估本身的具体比较和主张-评估-证据论证结构。

第一次适用性检查中，效度框架在行为研究人员和设计科学家参加的区域信息系统研讨会上进行了介绍；我们的重点是初步评估效用并在聘请最后一批设计科学专家之前进行任何调整。15 名参与者接受了使用效度框架的培训，并负责将四个效度定义和四个设计科学评估组织到框架中并回答一份开放式调查。结果表明，该框架对设计科学家和其他研究人员很有用。例如，该框架被描述为“一个发人深省、令人耳目一新的视角”。其他评论指出，该框架相当复杂，尤其是对于设计科学经验有限的研究人员而言。作为回应，我们增加了一个流程来选择在研究项目中使用哪些效度类型，并简化了框架。

将反馈纳入框架后（最终版本如图 2 所示），我们在领先的设计科学家中进行了第二次适用性检查，以评估框架的重要性（标准主张）、可及性（环境主张）和适用性（环境主张）。

这 11 位参与者都是设计科学家，拥有不同程度的博士后学术经验（平均 13 年）。对于拥有 Google Scholar 帐户的人（9 位参与者），他们的平均引用次数为 4,981。检查他们的论文发现，36% 的论文引用了 Prat 等人（2015 年）或 Venable 等人（2016 年）。适用性检查是在线进行的，参与者来自不同地区、研究领域和设计科学子社区。这些参与者采用了各种方法，考虑了不同的研究主题和分析类型。他们的研究涵盖了广泛的主题，包括数据和知识建模、分析和机器学习建模、数据科学、业务流程变更、模拟和设计理论。我们特别邀请了广泛领域的参与者，包括整个设计科学领域的参与者。参与者是 WITS、DESRIST 和其他设计科学会

议的定期撰稿人。适用性检查的各个阶段如下所述。

准备： 参与者被要求描述他们在正在进行的研究或最近完成的项目中使用的验证流程的例子。这些现有的流程是专家心理模型的一部分，也是参与者评估我们框架的基准。

总体而言，参与者对构件评估和验证的必要性表现出一致的理解，但对与他们的评估相符的效度类型的理解有限，并且缺乏描述其验证的词汇。在很多情况下，效度类型都是用心理测量术语（例如，构念效度）或与效用和功效相关的术语（例如，概念证明、价值证明）来讨论的。大多数参与者关注的是执行评估的过程（例如，现场实验、实验室实验），而不是验证在支持他们构件主张方面的作用。

设计科学效度框架简介： 随后，参与者参加了一个小时的在线会议，作者团队在会议期间介绍了效度框架。⁷ 在解释了框架和扩展示例之后，要求参与者使用我们的流程来确定他们项目中的效度类型。

反馈： 完成任务后，参与者填写了一份调查问卷，其中包含六个开放式问题和一组关于构件有用性和使用意图的问题，采用 Venkatesh 等人（2003）的李克特量表。定性和定量的回应表明，参与者认为效度框架对于满足社区的既定需求很重要、易于理解也合适（Rosemann 和 Vessey 2008），并相信它对他们自己的研究有用。

Rosemann 和 Vessey（2008）将重要性定义为“通过解决现实世界的问题来满足实践需求

的研究……这样它就可以作为提供解决方案的起点”（第 3 页）。我们认为评估这一方面的过程是标准效度的一个例子。在我们的环境中，设计科学家是现实世界中的从业者，他们自己需要验证有关构件的知识主张。可及性是一种标准效度，它“涵盖了研究是否易于理解、可读，以及是否注重结果而不是研究过程”（第 3 页）。最后，适用性被定义为研究能够“[满足]实践需求”的程度（第 3 页），我们认为这意味着研究人员认为该框架适合目标环境的程度。鉴于对这三项评估的评估是在与现实世界相似的环境中进行的，因此它们都涉及模型效度和环境效度。

参与者一致认为，该框架对于澄清和构建日益复杂的验证方法、标准和策略要求非常重要。诸如“它支持系统化的验证方法”和“这样的框架可以为研究人员提供一个公认的标准，让他们在尝试验证自己的设计科学成果时可以参考”等评论证明了这一重要性。其他评论也承认，该框架降低了复杂性，将知识主张与验证策略联系起来，并提供了有助于阐述效度主张的细节。所有这些评论都支持我们的标准主张。

尽管只有短暂的框架介绍，可及性得到了以下评论的证实：“我在项目早期就了解了您的效度框架，这是一件好事。”一位参与者发现该框架可以立即可及，并表示：“我可以使用框架……该框架验证了我的评估足够好，但我应该在评估中为审阅者详细说明一些细节。”

最后一项适用性前提，即适用性，也反映在评论中：“作为 DSR 学者和作者，该框架可

⁷该演示文稿可在透明材料<https://osf.io/ca6vg/>上找到。

帮助我尽早规划我的工作验证。……作为审稿人和编辑，我将能够指出对 DSR 效度的共同理解。在评估 DSR 论文的价值时，我可以使用该框架来确定优势和改进潜力。或者促使作者分享他们从论文中排除但与建立其效度相关的 DSR 工作方面。”

关于该框架的未来影响，人们提出了两个担忧。首先，该框架可能被用来抑制设计科学的出版，因为“审稿人可能对需要哪些效度类型意见不一”。与会者一致认为，这并不是框架本身的缺陷。将该框架作为额外评估的论据是不恰当的。如上所述，只有当所进行的评估与所提出的知识主张不一致或审稿小组认为论文对现有知识主张的贡献不足时，才有理由进行额外评估。在后一种情况下，审稿人提出的任何额外评估都应适用于额外的知识主张。

第二个担忧是，由于对效度程序的要求不断增加，制作和审查设计科学工作的成本和时间要求可能会增加。我们更新了我们的初始框架，以更好地解释知识主张和应用该框架的过程，以防止此类框架的使用。

参与者根据 7 点李克特量表回答了该框架的实用性 ($\mu = 5.64$; $\sigma = 1.04$)，这是设计科学中常见的适用性检查问题（例如，Li 等人，2020 年；Lukyanenko 等人，2019 年）。他们还被问及相对于现有流程使用该框架的意图 ($\mu = 6.15$; $\sigma = .92$)，这是另一项标准主张，在目标设定中评估时，也可作为环境主张。对于实用性，一个异常指标反映了第二个担忧，即该框架是否会（至少在最初）减慢研究任务的速度。没有这个指标，实用性显著增加（从 $\mu = 5.64$ 到 $\mu = 6.14$; $\sigma = 1.03$ ）。

一位担心该框架可能难以使用的参与者也打算使用它，他表示“这样的框架的存在可以

为研究人员提供一个公认的标准，让他们在尝试验证自己的设计科学成果 [主张] 时可以参考。”所有参与者都表示，一旦该框架可用，他们就会使用它。因此，我们得出结论，适用性检查确立了模型效度，并增加了生态效度的初步证据。

主张 2：框架完整性的标准主张

我们验证了效度框架在代表已发表的设计科学文章中的现有评估方面是完整的这一说法。由于该框架是一个模型，它应该能够代表这些评估，从而产生模型效度。我们咨询了两个来源，以确定要抽样的相关论文群体。首先，我们使用了 Prat 等人(2015)从信息系统协会 (AIS) 高级学者八种期刊 (2004 年 4 月至 2013 年 12 月) 中确定的 121 篇设计科学论文。其次，我们分析了来自 AIS 的另外 1,233 篇候选论文

2014 年至 2017 年的八篇论文。然后，我们根据 Prat 等人 (2015 年) 的纳入方案手动编码了这 1,233 篇论文。第三作者将 86 篇论文编码为设计科学，并由第二作者进行交叉核对。所有论文均为设计科学论文，符合率 100%。Prat 等人 (2015 年) 最初的 121 篇论文加上我们的 86 篇论文，共计 207 篇论文发表于 2004 年 4 月至 2017 年 12 月期间。

为了捕获遵循设计科学方法但未明确使用“设计科学”一词的文章，我们根据设计科学关键词列表以及顶级设计科学论文的引用创建了特征集。然后，我们根据 Larsen 等人(2019)描述的过程，训练了一个机器学习模型来区分全套手稿中的相关手稿。我们结合使用机器学习和人工评估，将覆盖年份从 1994 年扩大到 2019 年，并将来源扩展到包括 AIS Basket of Eight、决策支持系统、ACM Transactions on Information Systems

表 6. 效度定义 (阴影单元格表示“不适用”)

姓名	部分	输出	构件	船用运输系统
设计科学效度				
标准效度				
标准效度			19	
预测效度			3	
同时效度			4	
标准特征效度				
实例效度	0	0	5	
理论效度	0	0	0	
模型效度	0	0	0	
方法效度	0	0	10	
因果效度				
因果效能效度	24			
因果特征效度	7	0	1	
环境效度				
外部效度				7
生态效度				1

以及信息系统和技术设计科学研究国际会议 (DESRI) 和国际信息系统会议 (ICIS) 的论文集。在对最后四个来源进行下采样后，我们最终得到了 527 篇设计科学论文。

我们从语料库中随机抽取了 32 篇实证文章，由第四和第六作者独立编码，确定了 79 个评价描述。然后，第一和第三作者使用该框架独立编码每个评价描述。编码员在 76.3% 的情况下意见一致，Cohen's kappa 为 0.703。分歧通常涉及原作者想要声明的内容，而不是如何解释效度框架。通过这项练习，我

们得出结论，可以使用我们框架中的效度类型对所审查论文中发现的评价进行分类，从而证明模型效度。但是，两种效度类型，理论效度和模型效度，在我们所抽样的论文中并没有使用 (表6)。

主张 3：代表权的标准和环境主张

共享学科概念的形式化加速了科学进步 (美国国家科学院 2022)。其他学科也开发和验证构件，效度框架必须能够代表用于评估这些构件的效度。

第三个主张是，效度框架在表示设计科学效度类型方面是完整的，不仅适用于信息系统，也适用于行为科学、工程和医学——当这些学科验证其构件的主张时。通过评估标准主张，即框架在表示设计科学效度定义方面是完整的，我们建立了模型效度。通过表明我们的框架适用于其他学科，我们建立了外部效度。

第一个任务是确定文献中已经提出的特定设计科学效度类型。没有发现大量现有的通用效度类型集。因此，我们建立了这样一个数据集，以进一步细化效度类型的子集。我们首先从社会科学、工程和计算机科学以及医学等各个领域识别和收集效度类型和定义。这是第一作者和一组研究助理在三年内编写的。最初资料来源是包含效度类型集文档，例如教育和心理测试的标准 (例如，美国教育研究协会等，2014 年)。最大的效度名称集合，168 个，由 Newton 和 Shaw 提供 (2014 年)。其中绝大多数都放在一个大表中，以展示效度类型的难处理性。然而，许多并没有出现在任何可用的数据库中，也无法使用搜索引擎找到。

第二项任务是获得已识别效度类型的定义。我们主要依靠科学书籍和文章。对于每个来

源，进一步检查包含效度的部分以确定更多的候选效度类型。候选效度是作者主张为效度的概念，或者未声明为效度但列出效度类型的概念（例如，*世俗现实主义与生态效度*密切相关）。我们没有质疑作者的陈述。在每种情况下，我们都记录了新的候选效度类型。总共有 2,418 种候选效度类型来自大约 7,500 个手动检查的来源。在检查每个类别集以查找同音异义词（名称相同但含义不同）之前，对它们按名称进行分类，从而产生 418 种不同的效度类型。该文献综述用于生成设计科学效度框架的第一个版本。

在第三项任务中，我们检查了 1994 年至 2017 年在 AIS Basket of Eight 期刊 (Lowry 等人, 2013 年) 上发表的所有文章。通过应用 216 个代表效度、可靠性和相关概念（如普遍性）的正则表达式查询，共分析了 6,083 篇文章，共产生了 73,365 个句子。句子按命中次数排序，其中 9,707 个句子在第一作者手动检查的正则表达式上包含两个以上的命中。此分析仅产生了 23 种额外的效度类型，总共 441 种候选效度类型，这表明原始过程已经非常彻底。我们从进一步评估中删除了任何不常用的候选效度类型，因为我们无法从不同来源找到五个定义。在此步骤中，删除了 158 种候选效度类型。

在第四项任务中，第一和第二作者分别阅读了每种效度类型的五个定义，并选择了一两个代表其他定义所表达的整体方面的定义。讨论（并记录）了这些决定，并解决了分歧，以就代表效度类型的一两个定义达成一致。

由于存在高度相似甚至相同的效度类型，因此计算评分者间信度指标并不合适。在此步骤中，11 种候选效度类型被淘汰，因为它们不符合我们对研究效度的定义（例如，基于法律的效度）或因为没有出现明确的定义。⁸

对于第五项任务，在指定最终一组效度类型后，第一和第二作者遵循我们既定的设计科学效度定义，并独立将所有效度类型编码为“设计科学效度类型”或“其他效度类型”。编码员达成 90.4% 的一致性，Cohen's kappa 为 0.79。编码员讨论并解决了任何分歧。几乎所有分歧都源于功效效度类型，其中一名编码员采用了更具包容性的解释。总共发现了 79 种设计科学效度类型。

我们从设计科学效度类型中移除了非叶效度类型，这些类型结合了多种其他效度类型。移除这些效度类型后，我们框架中仍有 70 个定义可供分类。其中，23 个（33%）来自行为科学机构，37 个（53%）来自工程和计算机科学，10 个（14%）来自医学；22 个（31%）来自 IS 期刊样本。

第一作者和第三作者独立将所有效度定义编码到框架中，并在 79.2% 的案例中达成一致 (Cohen's kappa = .731)，其中包括编码为不同意的部分一致案例（例如，标准功效效度与预测效度）。分歧得到了讨论和解决。所有定义都符合框架的效度类别（表 7 左栏）。⁹

⁸ 每次删除的理由均可在透明材料中找到。

⁹ 注意：效度指标已包含在分析中，但由于它们未采用主张，因此它们本身不是效度。因此，我们从分类中排除了以下指标：准确度、曲线下面积、完整性、正确拒

绝、检测率、F1 分数、后果、误报、错误发现比例、假阴性、假阴性率、假遗漏率、假阳性、假阳性率、命中、命中率、马修斯相关系数、遗漏、遗漏率、阴性预测值、阳性预测值、精确度、召回率、灵敏度、特异性、真阴性、真阴性率、真阳性、真阳性率。

表 7. 效度定义编码	
设计科学效度	文献中的效度
标准效度	
标准效度	绝对效度、标准群体效度、标准效度、标准导向效度、标准相关效度、决策效度、诊断效度、判别效度、经验效度、低阶效度、操作效度、实用效度、程序效度、事后效度、复制效度、回溯效度、应用效度
预测效度	预测标准效度、预测效度 (2)* ， 预测标准相关效度
同时效度	同时标准效度、同时标准相关效度、同时效度、横断面效度、相对效度
标准特征效度	
实例效度	观察效度、物理效度
理论效度	病因效度、理论效度、实例效度
模型效度	概念模型效度、功能效度、结构效度、语义效度
方法效度	算法效度、一致性
因果效度	
因果效能效度	
因果特征效度	
环境效度	
外部效度	实用效度
生态效度	行为效度、生态效度

注意： *两个略有不同的预测效度版本引用相同的效度类型。

因此，评估从标准主张和上下文主张两个方面为主张 3 的效度提供了强有力的证据，加强了模型效度，并提供了外部效度。它不支持框架中包含两种因果效度的必要性。然而，我们在第 5.2 节中的评估已经确定了因果效度类型的必要性。现有的效度类型定义中缺

乏对因果效度的关注，这表明有机会为设计科学效度做出贡献，进而为评估做出贡献。从这个意义上说，它支持 Peffers 等人 (2008) 提出的设计科学模板的必要性。

主张 4：框架是简约的因果主张

设计科学效度框架是迭代开发的，这在设计科学中很常见。由于我们希望保持包容性，

该框架的初始版本比本文中报告的版本更为广泛。例如，初始框架包含一种称为“需求效度”的效度类型 - 一种在将构件与需求文档或用户的期望和体验进行比较时使用的标准特征效度类型。但是，需求不是捕捉对科学和社会贡献的设计科学构件，我们无法明确将需求效度归类为一种独特的效度类型。因此，我们进行了*因果特征效度*评估，以检查对这种效度类型的必要性（支持以下说法：没有这种效度类型的框架是简约的，并且不需要更多的组件）。

引入需求效度是为了解决针对明确需求（例如功能、易用性、形式）和隐含需求（例如需求、目标或对类似构件的体验）的知识主张。*需求*是软件开发中产生的常见构件。尽管如此，它并没有出现在设计科学构件类型的常见列表中（例如，Gregor and Hevner 2013；March and Smith 1995）。然而，我们最初认为它是我们框架的主要参考实体，偏向于更包容和保守。随着我们对效度领域的理解随着时间的推移而扩大，我们开始质疑*需求是否* 是需要的。我们选择通过正式主张“*需求是框架完整的必要效度类型*”来解决此问题。

为了验证这一主张，我们选择了一种基于反事实推理的评估方法。第一作者和第三作者合作编码了两个效度定义，这两个定义之前在原始框架中被编码为“要求效度”。首先，我们检查了“应用效度”，定义为模拟模型是否符合其目的和要求；即模型产生反映某些外部构件或社会技术系统的输出的可能性。基于我们对框架的更好理解，这个效度实例被重新编码为*标准功效效度*。第二个现有定义是“语义效度”，部分基于类别定义的适当性来定义。我们意识到参考实体是语义评估者对等效模型或社会技术系统类别的理解，这表明这是一种*模型效度*。在检查最初被归类为“要求效度”的七个设计科学评估时，

同样的逻辑变得更加清晰。在大多数情况下，初步评估侧重于“要求”或“满意度”和“可接受性”等术语。通常，原作者对评估的描述很差，而且在实际参考构件或社会技术系统方面也不清楚。在这次因果效度评估中，我们关注的是我们认为研究人员认为的参考构件。在七次评估中，我们发现了五次*模型效度案例*和两次*实例效度案例*。

因此，该评估表明，效度框架可以变得更加简约，而不会失去表征能力。因此，对*因果效度的反事实评估*表明，框架中被删除的部分与框架的性能没有因果关系。其他效度类型都是围绕公认的构件类型构建的，并且对于对证据进行分类是必要的，这表明原始简约主张对框架其余部分的*因果特征效度*。因此，额外的因果效度评估不太可能能够删除其他部分。因此，认为没有必要进行进一步的迭代。

讨论与启示

本研究有助于提高设计科学的效度，并可将其扩展到其他应用。我们的工作具有理论和实践意义。

一般贡献

设计科学的一个关键部分是验证关于焦点构件的知识主张。本研究阐明了为什么需要验证程序，并提供了一个可以通过各种效度类型识别和实施验证程序的过程。

我们提出了设计科学效度框架。它将效度类型映射到知识主张的特征，帮助作者制定和传达他们的知识主张以及支持它们的证据。基于有史以来最大规模的效度文献审查，设计科学效度框架是一个用于验证有关构件的知识主张的综合框架。该框架为研究效度报告提供了标准词汇。它具有外部效度，因为它成功地代表了信息系统、行为科学、工程、

计算机科学和医学中所有已识别的与构件相关的效度证据。

该框架通过考虑效度类型层次结构的分支来指导研究人员识别有关构件的知识主张，从而加强设计科学项目的严谨性和贡献。通过提供验证知识主张的过程和术语，该框架可应用于任何类型的构件创建和评估。明确地将效度类型、评估过程和支持证据与知识主张联系起来，对于设计科学的研究人员和审稿人以及采用其他研究方法的研究人员和审稿人都很有用。因此，我们的框架有助于建立设计科学与其他类型研究之间的联系。明确阐述的知识主张也可用于扩展先前工作的贡献。例如，可以通过对构件提出因果和环境主张，或添加具有更新比较实体的额外标准主张来实现。

发现与启示

构建和评估设计科学效度框架产生了几值得注意的见解，证明了其价值。首先，设计科学家历来使用临时评估来支持关于构件本身效度的隐性主张。为了将效度作为设计科学严谨性的一个方面来集中和扩展，我们将这些评估转移到确定关于构件的明确知识主张的效度。这些主张将具有不同程度的支持证据，其充分性将由社区确定，并可能随着时间的推移受到质疑和改变。

其次，尽管效度的评估和类型在其他学科中已经得到充分确立和理解，但令人惊讶的是，设计科学效度的全面讨论却很少。已发表的设计科学使用了范围狭窄的效度概念，主要侧重于功效测量和特征效度类型，这表明效度尚未得到充分利用。在我们的文献回顾中，最常出现的术语是准确度、精确度、召回率、特异性、真阳性和假阳性（与混淆矩阵相关的概念）。事实上，这些是指标而不是效度类型，但用于建立功效效度。在所分析的出

版物样本中，这些效度指标最常与机器学习模型的评估有关，当然，这不代表设计科学的范围。

第三，设计科学中使用的一些效度类型是从其他学科借鉴而来的。然而，这不利于设计科学的整体评估方法。例如，心理测量学和计量经济学中的效度主要侧重于测量，而一些类型在设计科学中很有用（例如，内部效度支持实验中的因果关系主张）。然而，心理测量效度不足以支持设计科学项目中的一系列知识主张。

最后，正如我们在适用性检查中所确认的那样，对于创建构件的研究人员来说，一个主要挑战在于验证的不确定性以及倾向于隐性地执行验证活动，而对此类验证的底层结构知之甚少。效度框架为研究人员提供了一个结构化的模板和一组经过仔细解释的效度类型，这些效度类型具有明确的标准化定义。验证过程允许研究人员收集支持与设计构件的知识主张明确相关的效度类型的证据。

实用建议

总结我们论文的论点和发现，我们为有兴趣开发对科学和社会有贡献的构件的研究人员提出以下建议。

明确陈述知识主张。为了获得高水平的实用性和科学可重复性，研究人员应该阐明关于其构件的知识主张。这有助于构建研究贡献并指导验证，因为验证取决于清晰的知识主张。此外，明确的知识主张支持知识的积累。

鉴于设计知识可以通过对构件的迭代细化而发展（Tuunanen 等人，2024），声明可以在迭代过程中的各个阶段出现。例如，某些主张可以在构件开发之前提出，而其他主张可以在将构件部署到某些环境中并观察结果

之后提出。但是，后者的主张在经过适当的验证程序（可能是在构件的后续迭代中）之前不应被视为已验证。重要的是，知识主张的正式评估独立于生成主张的过程，这要求主张必须明确陈述。

提出与预期贡献相称的主张。设计科学效率度框架不应被用来证明过度验证的合理性。陈述关于构件的每一个主张通常是不可能/没有必要的，而且并非所有主张都可以接受验证（例如，由于难以进行比较或获取合适的比较实体）。需要多少证据取决于问题的环境和具体特征。

研究人员应就该构件提出至少一个标准主张，努力与最先进的构件或流程进行比较。如果构件特别新颖，以至于人们对其效率度或可部署的其他环境（除原始环境外）了解甚少，这可能就足够了。除此之外，因果和环境主张加强了研究贡献，因为它们深化了设计知识，并帮助从业者可靠、安全地将研究结果应用于各种环境。

确保每项知识主张都经过验证。如果对构件提出了原始主张，则应对其进行验证。例如，如果构件的某些组件被认为会导致特定结果，则建立因果效率度是适当的。虽然单一验证不能证明知识主张，但验证主张会增加产生可靠设计知识的可能性。

至于需要多少评估才能确定主张的效率度，Galison 和达戈斯蒂诺（1987 年）提出令人信服的论据，即证据的充分性是社区共识的问题。所需的验证活动数量通常不是基于特定规则，而是基于对“有说服力的论据，即‘在法庭上站得住脚的论据’”的汇集的共识（Galison 和 D'Agostino 1987，第 227 页）。研究人员和审查团队在审查过程中达成了这样的共识。重要的是，所采取的验证程序适合验证论文中的知识主张。

在整个构件历程中验证主张。许多构件通过实验、修修补补或逐步改进等迭代过程而出现。一些项目涉及多个阶段，并在这些阶段产生不同的构件（例如概念模型、系统原型、测试版、最终生产系统）（Tuunanen 等人，2024）。可以在整个过程中对最终构件做出知识主张。对这些主张进行形成性验证有助于进一步完善和改进构件。“中间”验证可以是非正式的，例如使用潜在用户的便利样本或易于获得的标准。然而，为了确保最终构件能够可靠地为科学和实践做出贡献，关于构件最终或公开版本的主张应经过严格的总结性验证。

确保使用适当的效率度类型。随着设计科学效率度框架的建立，研究人员可以在验证其知识主张时参考该框架。该框架将各种效率度实践组织成一个连贯的参考系统。它显示了哪些效率度类型适合每种主张类型，并建议了适合这些效率度类型的比较实体和比较程序。随着社区继续应用和完善这些效率度类型，它们的稳健性有望提高，为研究人员开展研究奠定更坚实的基础。

设计科学效率度框架为未来的研究提供了机会。首先，该框架是可扩展的，这意味着它可以容纳研究界可能提出的其他效率度子类型（例如，根据模型类型进一步细化模型效率度）。其次，可以更好地追踪设计科学中的验证模式，并识别差距和机会（例如，需要更多的环境主张）。一个相关的可能性是通过确保系统地捕获其各自效率度类型的适当验证实践来改进验证程序。最后，研究人员可以调查该框架在设计科学项目中的应用，并报告与使用该框架对设计科学成熟度及其与其他研究传统的整合的实用性和长期影响相关的结果。

结论

这项研究定义了设计科学的效度，并提出了设计科学效度框架及其使用流程。该框架源自对效度文献的广泛审查和分析，将隐性和显性效度类型识别并组织成三个主要类别：*标准*、*因果*和*环境*。效度类别旨在帮助研究人员严格获取和呈现其知识主张的证据。将知识主张映射到效度框架上应该可以为学者提供支持，并有助于将与信息系统构件相关的科学知识联系起来。该框架本身通过评估其自身的知识主张并为支持这些主张的相关效度类型提供支持证据来验证。

致谢

作者感谢高级编辑 Balaji Padmanabhan、副编辑 Stefan Seidel 和匿名审稿人对我们论文先前版本所做的工作。我们感谢 HICSS 和 ECIS 研讨会以及科罗拉多大学、亚利桑那大学、亚利桑那州立大学和悉尼大学的参与者提供的反馈。

资金。 本研究部分由美国国立卫生研究院向 Kai R. Larsen 提供的资金 [拨款3U24AG052175-08 S1]以及加拿大自然科学与工程委员会致 Jeffrey Parsons的资助 [Grant RGPIN-2020-04916]。

参考书目

Abbasi, A., Albrecht, C., Vance, A., and Hansen, J. 2012. “Metafraud: A Meta-Learning Framework for Detecting Financial Fraud,” *MIS Quarterly*, pp. 1293-1327.

Abbasi, A., and Chen, H. 2008. “CyberGate: A Design Framework and System for Text Analysis of Computer-Mediated Communication,” *MIS Quarterly* (32:4), pp. 1-30.

Abbasi, A., Zhou, Y., Deng, S., and Zhang, P. 2018. “Text Analytics to Support Sense-Making in Social Media: A Language-Action Perspective,” *MIS Quarterly* (42:2), pp. 1-38.

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational, and Psychological Testing (US). 2014. *Standards for Educational and Psychological Testing*, American Educational Research Assn.

APA. 2020. “Validity,” American Psychological Association. (<https://dictionary.apa.org/validity>).

Arazy, O., Kumar, N., and Shapira, B. 2010. “A Theory-Driven Design Framework for Social Recommender Systems,” *Journal of the Association for Information Systems* (11:9), pp. 455-490.

Avdiji, H., and Winter, R. 2019. “Knowledge Gaps in Design Science Research,” in *DESRIST*, pp. 1-9.

Avoine, G., Bingol, M. A., Kardas, S., Lauradoux, C., and Martin, B. 2009. “A Formal Framework for Cryptanalyzing Rfid Distance Bounding Protocols,” This Work Is Partially Funded by FP7-Project ICE under the Grant Agreement (206546), Citeseer.

Baskerville, R., Kaul, M., and Storey, V. C. 2015. “Genres of Inquiry in Design-Science Research: Justification and Evaluation of Knowledge Production,” *MIS Quarterly* (39:3), pp. 541-564.

Boudreau, M.-C., Gefen, D., and Straub, D. W. 2001. “Validation in Information Systems Research: A State-of-the-Art As

- essment,” *MIS Quarterly* (25:1), pp. 1-16.
- vom Brocke, J., Gau, M., and Mädche, A. 2021. “Journaling the Design Science Research Process. Transparency About the Making of Design Knowledge,” in *DESIRIS T 2021*, Springer, pp. 131-136.
- vom Brocke, J., Hevner, A. R., and Maedche, A. 2020. *Design Science Research: Cases*, Berlin / Heidelberg: Springer.
- vom Brocke, J., Winter, R., Hevner, A., and Maedche, A. 2020. “Accumulation and Evolution of Design Knowledge in Design Science Research: A Journey through Time and Space,” *Journal of the Association for Information Systems*.
- Burton-Jones, A., Boh, W. F., Oborn, E., and Padmanabhan, B. 2021. “Editor’s Comments: Advancing Research Transparency at *MIS Quarterly*: A Pluralistic Approach,” *MIS Quarterly* (45:2), pp. iii-xviii.
- Carter, J. W. 2019. *Aristotle on Earlier Greek Psychology: The Science of Soul*, Cambridge, UK: Cambridge University Press.
- Chan, E. K. 2014. “Standards and Guidelines for Validation Practices: Development and Evaluation of Measurement Instruments,” in *Validity and Validation in Social, Behavioral, and Health Sciences*, Springer, pp. 9-24.
- Chatterjee, S., Sarker, S., Lee, M. J., Xiao, X., and Elbanna, A. 2021. “A Possible Conceptualization of the Information Systems (IS) Artifact: A General Systems Theory Perspective 1,” *Information Systems Journal* (31:4), Wiley Online Library, pp. 550-578.
- Chowdary, G. J., and Kanhangad, V. 2022. “A Dual-Branch Network for Diagnosis of Thorax Diseases From Chest X-Rays,” *IEEE Journal of Biomedical and Health Informatics* (26:12), IEEE, pp. 6081-6092.
- Cohen, L., Manion, L., and Morrison, K. 2013. *Research Methods in Education*, London England: Routledge.
- Collier-Reed, B., and Ingerman, Å. 2013. “Phenomenography: From Critical Aspects to Knowledge Claim,” in *Theory and Method in Higher Education Research* (Vol. 9), Emerald Group Publishing Limited, pp. 243-260.
- Collins, J., Hall, N., and Paul, L. A. 2004. *Causation and Counterfactuals*, Cambridge, MA: MIT Press.
- Cook, T. D., and Campbell, D. T. 1979. *Quasi-Experimentation: Design & Analysis Issues for Field Settings*, Chicago: Rand McNally College Pub. Co.
- Creswell, J. W., and Miller, D. L. 2000. “Determining Validity in Qualitative Inquiry,” *Theory into Practice* (39:3), Taylor & Francis, pp. 124-130.
- Ethayarajah, K., and Jurafsky, D. 2020. “Utility Is in the Eye of the User: A Critique of NLP Leaderboards,” *ArXiv Preprint ArXiv:2009.13888*.
- Etudo, U., Yoon, V., and Liu, D. 2017. “Financial Concept Element Mapper (FinCEM) for XBRL Interoperability: Utilizing the M3 Plus Method,” *Decision Support Systems* (98), Elsevier, pp. 36-48.
- Fan, W., Liu, X., Lu, P., and Tian, C. 2018. *Catching Numeric Inconsistencies in Graphs*, presented at the Proceedings

of the 2018 International Conference on Management of Data, pp. 381-393.

Galison, P. L., and D'Agostino, S. 1987. *How Experiments End*, Chicago, IL: University of Chicago Press Chicago.

Germonprez, M., Hovorka, D. S., and Collopy, F. 2007. "A Theory of Tailorable Technology Design," *Journal of the Association for Information Systems* (8:6), pp. 351-367.

Gonzalez-Huerta, J., Boubaker, A., and Mili, H. 2017. "A Business Process Re-Engineering Approach to Transform BPMN Models to Software Artifacts," in *E-Technologies*, pp. 170-184.

Gregor, S., and Hevner, A. R. 2013. "Positioning and Presenting Design Science Research for Maximum Impact," *MIS Quarterly* (37:2), pp. 337-355.

Gregor, S., and Jones, D. 2007. "The Anatomy of Design Theory," *Journal of the Association for Information Systems* (8:5), pp. 312-335.

Gregor, S., Kruse, L. C., and Seidel, S. 2020. "The Anatomy of a Design Principle," *Journal of the Association for Information Systems* (21:6), pp. 1622-1652.

Gregory, R. W., and Muntermann, J. 2014. "Research Note-Heuristic Theorizing: Proactively Generating Design Theories," *Information Systems Research* (25:3), *INFORMS*, pp. 639-653.

Guba, E. G., and Lincoln, Y. S. 1994. "Competing Paradigms in Qualitative Research," *Handbook of Qualitative Research* (2:163-194), California, Sage Publications, p. 105.

Gumport, P. J. 2007. *Sociology of Higher Education: Contributions and Their Contexts*, Washington, DC: Johns Hopkins University Press+ ORM.

Hevner, A., March, S., Park, J., and Ram, S. 2004. "Design Science in Information Systems Research," *MIS Quarterly* (28:1), pp. 75-105.

Hevner, A., Parsons, J., Brendel, A. B., Lukyanenko, R., Tiefenbeck, V., Tremblay, M. C., and vom Brocke, J. 2024. "Transparency in Design Science Research," *Decision Support Systems* (182:1), pp. 1-19.

Hoyningen-Huene, P. 2013. *Systematicity: The Nature of Science*, Oxford, UK: Oxford University Press.

Iivari, J. 2015. "Distinguishing and Contrasting Two Strategies for Design Science Research," *European Journal of Information Systems* (24:1), Springer, pp. 107-115.

Iivari, J., Rotvit Perlt Hansen, M., and Haj-Bolouri, A. 2021. "A Proposal for Minimum Reusability Evaluation of Design Principles," *European Journal of Information Systems* (30:3), Taylor & Francis, pp. 286-303.

Johannesson, P., and Perjons, E. 2014. *An Introduction to Design Science*, Berlin / Heidelberg: Springer.

Koornneef, H., Verhagen, W. J., and Curran, R. 2020. "A Decision Support Framework and Prototype for Aircraft Dispatch Assessment," *Decision Support Systems* (135), Elsevier, p. 113338.

Kuechler, W., and Vaishnavi, V. 2012. "A Framework for Theory Development in

- Design Science Research: Multiple Perspectives.” *Journal of the Association for Information Systems* (13:6), pp. 395-423.
- Larsen, K., and Bong, C. H. 2016. “A Tool for Addressing Construct Identity in Literature Reviews and Meta-Analyses,” *MIS Quarterly* (40:3), pp. 1-23.
- Larsen, K. R., and Becker, D. S. 2020. *Automated Machine Learning for Business: An Introduction to Accurate, Easy, and Fast Analytics*, New York NY: Oxford, U K.
- Larsen, K. R., Hovorka, D., Dennis, A., and West, J. D. 2019. “Understanding the Elephant: The Discourse Approach to Boundary Identification and Corpus Construction for Theory Review Articles,” *Journal of the Association for Information Systems* (20:7), p. 15.
- Li, J., Larsen, K. R., and Abbasi, A. 2020. “TheoryOn: A Design Framework and System For Unlocking Behavioral Knowledge Through Ontology Learning,” *MIS Quarterly*, pp. 1-55.
- Lincoln, Y. S., and Guba, E. G. 1985. *Naturalistic Inquiry* (Vol. 75), Thousand Oaks, CA: SAGE Publications.
- Lowry, P. B., Gaskin, J., Humpherys, S. L., Moody, G. D., Galletta, D. F., Barlow, J. B., and Wilson, D. W. 2013. “Evaluating Journal Quality and the Association for Information Systems Senior Scholars’ Journal Basket via Bibliometric Measures: Do Expert Journal Assessments Add Value?,” *MIS Quarterly*, pp. 993-1012.
- Lukyanenko, R., and Parsons, J. 2020. “Design Theory Indeterminacy: What Is It, How Can It Be Reduced, and Why Did the Polar Bear Drown?,” *Journal of the Association for Information Systems* (21:5), pp. 1-30.
- Lukyanenko, R., Parsons, J., Wiersma, Y., and Maddah, M. 2019. “Expecting the Unexpected: Effects of Data Collection Design Choices on the Quality of Crowdsourced User-Generated Content,” *MIS Quarterly* (43:2), pp. 634-647.
- Mandviwalla, M. 2015. “Generating and Justifying Design Theory,” *Journal of the Association for Information Systems* (16:5), p. 314.
- March, S. T., and Smith, G. F. 1995. “Design and Natural Science Research on Information Technology,” *Decision Support Systems* (15:4), pp. 251-266.
- Matadamas-Hernández, J., Román-Alonso, G., Rojas-González, F., Castro-García, M. A., Boukerche, A., Aguilar-Cornejo, M., and Cordero-Sánchez, S. 2012. “Parallel Simulation of Pore Networks Using Multicore Cpus,” *IEEE Transactions on Computers* (63:6), IEEE, pp. 1513-1525.
- Maxwell, J. 1992. “Understanding and Validity in Qualitative Research,” *Harvard Educational Review* (62:3), Harvard Education Publishing Group, pp. 279-301.
- Meth, H., Mueller, B., and Maedche, A. 2015. “Designing a Requirement Mining System,” *Journal of the Association for Information Systems* (16:9), p. 2.
- Moules, N. J., McCaffrey, J., and Field, J. 2015. *Conducting Hermeneutic Research: From Philosophy to Practice*, New York NY: Peter Lang Publishing.

- National Academies of Sciences, E., and Medicine. 2022. *Ontologies in the Behavioral Sciences: Accelerating Research and the Spread of Knowledge: Digest Version*, Washington, DC: The National Academies Press.
- Newell, A. 1975. "A Tutorial on Speech Understanding Systems," *Speech Recognition*, pp. 4-54.
- Newton, P., and Shaw, S. 2014. *Validity in Educational and Psychological Assessment*, Hoboken, NJ: Sage.
- Nunamaker, J. F., Chen, M., and Purdin, T. D. 1991. "Systems Development in Information Systems Research," *Journal of Management Information Systems* (7:3), pp. 89-106.
- Onwuegbuzie, A. J., and Leech, N. L. 2007. "Validity and Qualitative Research: An Oxymoron?," *Quality & Quantity* (41:2), Springer, pp. 233-249.
- Pääkkönen, T., Kekäläinen, J., Keskustalo, H., Azzopardi, L., Maxwell, D., and Järvelin, K. 2017. "Validating Simulated Interaction for Retrieval Evaluation," *Information Retrieval Journal* (20), Springer, pp. 338-362.
- Padmanabhan, B., Fang, X., Sahoo, N., and Burton-Jones, A. 2022. "Machine Learning in Information Systems Research," *MIS Quarterly* (46:1), pp. iii-xix.
- Peffer, K., Tuunanen, T., and Niehaves, B. 2018. "Design Science Research Genes: Introduction to the Special Issue on Exemplars and Criteria for Applicable Design Science Research," *European Journal of Information Systems* (27:2), pp. 129-139.
- Peffer, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. 2007. "A Design Science Research Methodology for Information Systems Research," *Journal of Management Information Systems* (24:3), pp. 45-77.
- Piramuthu, S., and Doss, R. 2017. "On Sensor-Based Solutions for Simultaneous Presence of Multiple RFID Tags," *Decision Support Systems* (95), Elsevier, pp. 102-109.
- Prat, N., Comyn-Wattiau, I., and Akoka, J. 2015. "A Taxonomy of Evaluation Methods for Information Systems Artifacts," *Journal of Management Information Systems* (32:3), pp. 229-267.
- Ramakrishnan, M., Gregor, S., Shrestha, A., and Soar, J. 2023. "Design Principles for Platform-Enabled Knowledge Commons with an Expository Instantiation," *Journal of the Association for Information Systems* (24:5), pp. 1413-1438.
- Refsgaard, J. C., Van der Sluijs, J. P., Brown, J., and Van der Keur, P. 2006. "A Framework for Dealing with Uncertainty Due to Model Structure Error," *Advances in Water Resources* (29:11), Elsevier, pp. 1586-1597.
- Roberts, P., and Priest, H. 2006. "Reliability and Validity in Research," *Nursing Standard* (20:44), Royal College of Nursing Publishing Company (RCN), pp. 41-46.
- Rosemann, M., and Vessey, I. 2008. "Toward Improving the Relevance of Information Systems Research to Practice: The Role of Applicability Checks," *MIS Quarterly* (32:1), pp. 1-22.

- Salmon, W. C. 1998. *Causality and Explanation*, Oxford University Press.
- Sedrakyan, G., Poelmans, S., and Snoeck, M. 2017. "Assessing the Influence of Feedback-Inclusive Rapid Prototyping on Understanding the Semantics of Parallel UML Statecharts by Novice Modellers," *Information and Software Technology* (82), Elsevier, pp. 159-172.
- Sein, M., Henfridsson, O., Purao, S., Rossi, M., and Lindgren, R. 2011. "Action Design Research," *MIS Quarterly* (35:1), p. 37.
- Shultz, K. S., Riggs, M. L., and Kottke, J. L. 1998. "The Need for an Evolving Concept of Validity in Industrial and Personnel Psychology: Psychometric, Legal, and Emerging Issues," *Current Psychology* (17), Springer, pp. 265-286.
- Taylor, C. S. 2013. *Validity and Validation*, Series in Understanding Statistics, New York NY: Oxford University Press USA.
- Thomas, M. A., Li, Y., and Lee, A. S. 2022. "Generalizing the Information Systems Artifact," *Information Systems Research* (33:4), INFORMS, pp. 1452-1466.
- Tiefenbeck, V., Goette, L., Degen, K., Tasic, V., Fleisch, E., Lalive, R., and Staake, T. 2016. "Overcoming Saliency Bias: How Real-Time Feedback Fosters Resource Conservation," *Management Science* (64:3), pp. 1458-1476.
- Tuunanen, T., Winter, R., and Vom Brocke, J. 2024. "Dealing with Complexity in Design Science Research," *MIS Quarterly* (48:2), pp. 427-458.
- Umapathy, K., Purao, S., and Barton, R. R. 2008. "Designing Enterprise Integration Solutions: Effectively," *European Journal of Information Systems* (17:5), Springer, pp. 518-527.
- Valecha, R., Sharman, R., Rao, H. R., and Upadhyaya, S. 2013. "A Dispatch-Mediated Communication Model for Emergency Response Systems," *ACM Transactions on Management Information Systems (TMIS)* (4:1), ACM New York, NY, USA, pp. 1-25.
- Venable, J., Pries-Heje, J., and Baskerville, R. 2016. "FEDS: A Framework for Evaluation in Design Science Research," *European Journal of Information Systems* (25:1), pp. 77-89.
- Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. 2003. "User Acceptance of Information Technology: Toward a Unified View," *MIS Quarterly*, pp. 425-478.
- Weinhardt, C., Kloker, S., Hinz, O., and van der Aalst, W. M. 2020. "Citizen Science in Information Systems Research," *Business & Information Systems Engineering* (62), Springer, pp. 273-277.
- Winter, S., Berente, N., Howison, J., and Butler, B. 2014. "Beyond the Organizational 'Container': Conceptualizing 21st Century Sociotechnical Work," *Information and Organization* (24:4), pp. 250-269.
- Zaitsev, A., and Mankinen, S. 2022. "Designing Financial Education Applications for Development: Applying Action Design Research in Cambodian Countryside," *European Journal of Information Systems* (31:1), Taylor & Francis, pp. 91-111.

- Abbasi, A.、Albrecht, C.、Vance, A. 和 Hansen, J., 2012 年。“元欺诈：用于检测金融欺诈的元学习框架”，*MIS Quarterly*, 第 1293-1327 页。
- Abbasi, A. 和 Chen, H. 2008. “CyberGate：计算机中介通信文本分析的设计框架和系统。”*MIS Quarterly* (32:4), 第 1-30 页。
- Abbasi, A.、Zhou, Y.、Deng, S. 和 Zhang, P. 2018 年。“文本分析支持社交媒体中的意义建构：语言行动视角”，*MIS Quarterly* (42:2), 第 1-38 页。
- 美国教育研究协会、美国心理学会、国家教育测量委员会、教育与心理测试标准联合委员会（美国）。2014. 《教育与心理测试标准》，美国教育研究协会。
- APA. 2020 年。“效度”，美国心理学会。（<https://dictionary.apa.org/validity>）。
- Arazy, O.、Kumar, N. 和 Shapira, B. 2010. “社交推荐系统的理论驱动设计框架”，*信息系统协会期刊*(11:9), 第 455-490 页。
- Avdiji, H. 和 Winter, R. 2019. “设计科学研究中的知识差距”，载于 *DESIRIST*, 第 1-9 页。
- Avoine, G.、Bingol, MA、Kardas, S.、Laurado, C. 和 Martin, B. 2009. “用于加密分析 Rfid 距离边界协议的正式框架”，这项工作部分由 FP7-Project ICE 根据资助协议资助 (206546), Citeseer。
- Baskerville, R.、Kaul, M. 和 Storey, VC 2015. “设计科学研究中的探究类型：知识生产的论证和评估。”，*MIS Quarterly* (39:3), 第 541-564 页。
- Boudreau, M.-C.、Gefen, D. 和 Straub, DW 2001. “信息系统研究中的验证：最新评估”，*MIS Quarterly* (25:1), 第 1-16 页。
- vom Brocke, J.、Gau, M. 和 Mädche, A. 2021. “记录设计科学研究过程。设计知识形成的透明度”，载于 *DESIRIST 2021*, Springer, 第 131-136 页。
- vom Brocke, J.、Hevner, AR 和 Maedche, A. 2020. *设计科学研究：案例*, 柏林/海德堡：施普林格。
- vom Brocke, J.、Winter, R.、Hevner, A. 和 Maedche, A. 2020 年。“设计科学研究中设计知识的积累和演变：穿越时空的旅程”，*信息系统协会期刊*。
- Burton-Jones, A.、Boh, WF、Oborn, E. 和 Padmanabhan, B. 2021 年。“编者评语：推进 *MIS Quarterly* 的研究透明度：一种多元化方法”，*MIS Quarterly* (45:2), 第 iii-xviii 页。
- Carter, JW 2019. 《亚里士多德论早期希腊心理学：灵魂的科学》，英国剑桥：剑桥大学出版社。
- Chan, EK 2014. “验证实践的标准和指南：测量工具的开发和评估”，载于《社会、行为和健康科学中的效度和验证》，Springer, 第 9-24 页。
- Chatterjee, S.、Sarker, S.、Lee, MJ、Xiao, X. 和 Elbanna, A. 2021. “信息系统 (IS) 构件的可能概念化：一般系统理论视角 1”，*信息系统期刊*(31:4), 威利在线图书馆, 第 550-578 页。
- Chowdary, GJ 和 Kanhangad, V. 2022. “用于通过胸部 X 光诊断胸部疾病的双分支网络”，*IEEE 生物医学和健康信息学期刊*(26:12), IEEE, 第 6081-6092 页。
- Cohen, L., Manion, L. 和 Morrison, K. 2013. 《教育研究方法》，英国伦敦：劳特利奇。
- Collier-Reed, B. 和 Ingerman, Å. 2013. “现象描述学：从批判方面到知识主张”，*高等教育研究理论与方法* (第 9 卷), Emerald Group Publishing Limited, 第 243-260 页。

- Collins, J.、Hall, N. 和 Paul, LA 2004。《因果关系与反事实》，马萨诸塞州剑桥：麻省理工学院出版社。
- Cook, TD, 和 Campbell, DT 1979. *准实验：现场设置的设计与分析问题*，芝加哥：Rand McNally College Pub. Co.。
- Creswell, JW 和 Miller, DL 2000。“确定定性探究的效度”，*理论转化为实践* (39:3)，Taylor & Francis, 第 124-130 页。
- Ethayarajh, K. 和 Jurafsky, D. 2020 年。“效用在于用户的眼中：对 NLP 排行榜的批判”，*ArXiv 预印本 ArXiv:2009.13888*。
- Etudo, U.、Yoon, V. 和 Liu, D. 2017。“用于 XBRL 互操作性的财务概念元素映射器 (Fin CEM)：利用 M3 Plus 方法”，*决策支持系统*(98)，爱思唯尔，第 36-48 页。
- Fan, W.、Liu, X.、Lu, P. 和 Tian, C. 2018。《捕捉图表中的数字不一致》，发表于 *2018 年国际数据管理会议论文集*，第 381-393 页。
- Galison, PL 和 D'Agostino, S. 1987。《实验如何结束》，伊利诺斯州芝加哥：芝加哥大学出版社。
- Germonprez, M.、Hovorka, DS 和 Collopy, F., 2007 年。“可定制技术设计理论”，*信息系统协会杂志*(8:6)，第 351-367 页。
- Gonzalez-Huerta, J.、Boubaker, A. 和 Mili, H. 2017。“将 BPMN 模型转化为软件工件的业务流程再造方法”，载于 *E-Technologies*，第 170-184 页。
- Gregor, S. 和 Hevner, AR 2013。“定位和展示设计科学研究以实现最大影响”，*MIS Quarterly* (37:2)，第 337-355 页。
- Gregor, S. 和 Jones, D. 2007。“设计理论的剖析”，*信息系统协会期刊*(8:5)，第 312-335 页。
- Gregor, S.、Kruse, LC 和 Seidel, S. 2020。“设计原则的剖析”，*信息系统协会期刊*(21:6)，第 1622-1652 页。
- Gregory, RW 和 Muntermann, J. 2014 年。“研究报告—启发式理论：主动生成设计理论”，*信息系统研究* (25:3)，INFORMS, 第 639-653 页。
- Guba, EG 和 Lincoln, YS 1994。“定性研究中的竞争范式”，《*定性研究手册*》(2:163-194)，加利福尼亚，Sage 出版社，第 105 页。
- Gumpert, PJ 2007。《*高等教育社会学：贡献及其背景*》，华盛顿特区：约翰霍普金斯大学出版社 + ORM。
- Hevner, A.、March, S.、Park, J. 和 Ram, S. 2004。“信息系统研究中的设计科学，” *MIS Quarterly* (28:1)，第 75-105 页。
- Hevner, A.、Parsons, J.、Brendel, AB、Lukyanenko, R.、Tiefenbeck, V.、Tremblay, MC 和 vom Brocke, J. 2024。“设计科学研究的透明度”，*决策支持系统*(182:1)，第 1-19 页。
- Hoyningen-Huene, P. 2013. *系统性：科学的本质*，英国牛津：牛津大学出版社。
- Iivari, J. 2015。“区分和对比设计科学研究的两种策略”，*欧洲信息系统期刊* (24:1)，Springer, 第 107-115 页。
- Iivari, J.、Rotvit Perlt Hansen, M. 和 Haj-Bolouri, A. 2021。“设计原则最低可重用性评估提案”，《*欧洲信息系统期刊*》(30:3)，Taylor & Francis, 第 286-303 页。
- Johannesson, P. 和 Perjons, E. 2014。《*设计科学导论*》，柏林 / 海德堡：Springer。
- Koornneef, H.、Verhagen, WJ 和 Curran, R. 2020 年。“飞机调度评估的决策支持框架和原型”，*决策支持系统*(135)，爱思唯尔，第 113338 页。
- Kuechler, W. 和 Vaishnavi, V. 2012 年。“设计科学研究理论发展框架：多种视角。” *信息系统协会期刊*(13:6)，第 395-423 页。

- Larsen, K. 和 Bong, CH 2016. “文献综述和荟萃分析中解决构建身份问题的工具”, *MIS Quarterly* (40:3), 第 1-23 页。
- Larsen, KR 和 Becker, DS 2020. 《商业自动化机器学习: 准确、简单、快速分析简介》, 纽约: 英国牛津。
- Larsen, KR、Hovorka, D.、Dennis, A. 和 West, JD, 2019 年。“理解大象: 理论评论文章的边界识别和语料库构建的话语方法”, *信息系统协会期刊*(20:7), 第 15 页。
- Li, J.、Larsen, KR 和 Abbasi, A. 2020 年。“TheoryOn: 通过本体学习解锁行为知识的设计框架和系统”, *MIS Quarterly*, 第 1-55 页。
- Lincoln, YS 和 Guba, EG 1985. 《自然主义探究》(第 75 卷), 加利福尼亚州千橡市: SAGE 出版物。
- Lowry, PB、Gaskin, J.、Humpherys, SL、Mood, GD、Galletta, DF、Barlow, JB 和 Wilson, DW, 2013 年。“通过文献计量方法评估期刊质量和信息系统协会高级学者期刊篮子: 专家期刊评估是否增加了价值?”, *MIS Quarterly*, 第 993-1012 页。
- Lukyanenko, R. 和 Parsons, J. 2020 年。“设计理论的不确定性: 它是什么, 如何减少, 以及北极熊为什么会溺水?”, *信息系统协会期刊*(21:5), 第 1-30 页。
- Lukyanenko, R.、Parsons, J.、Wiersma, Y. 和 Maddah, M. 2019 年。“期待意外: 数据收集设计选择对众包用户生成内容质量的影响”, *MIS Quarterly* (43:2), 第 634-647 页。
- Mandviwalla, M. 2015. “生成和证明设计理论”, *信息系统协会期刊* (16:5), 第 314 页。
- March, ST 和 Smith, GF 1995. “信息技术设计与自然科学研究”, *决策支持系统*(15:4), 第 251-266 页。
- Matadamas-Hernández, J.、Román-Alonso, G.、Rojas-González, F.、Castro-García, M A、Boukerche, A.、Aguilar-Cornejo, M. 和 Cordero-Sánchez, S. 2012. “使用多核 CPU 的孔隙网络并行模拟”, *IEEE Transactions on Computers* (63:6), IEEE, 第 1513-1525 页。
- Maxwell, J. 1992. “定性研究中的理解和效度”, *哈佛教育评论* (62:3), 哈佛教育出版集团, 第 279-301 页。
- Meth, H.、Mueller, B. 和 Maedche, A., 2015 年。“设计需求挖掘系统”, *信息系统协会期刊*(16:9), 第 2 页。
- Moules, NJ、McCaffrey, J. 和 Field, J. 2015. 《进行诠释学研究: 从哲学到实践》, 纽约: Peter Lang Publishing。
- 美国国家科学、工程与医学院。2022 年。《行为科学中的本体论: 加速研究和知识传播: 文摘版》, 华盛顿特区: 美国国家科学院出版社。
- Newell, A. 1975. “语音理解系统教程”, *语音识别*, 第 4-54 页。
- Newton, P. 和 Shaw, S. 2014. 《教育与心理评估的效度》, 新泽西州霍博肯: Sage。
- Nunamaker, JF, Chen, M., 和 Purdin, TD 1991. “信息系统研究中的系统开发”, *管理信息系统期刊*(7:3), 第 89-106 页。
- Onwuegbuzie, AJ 和 Leech, NL 2007. “效度和定性研究: 矛盾修饰法?”, 《质量与数量》(41:2), Springer, 第 233-249 页。
- Pääkkönen, T.、Kekäläinen, J.、Keskustalo, H.、Azzopardi, L.、Maxwell, D. 和 Järvelin, K. 2017. “验证检索评估的模拟交互”, *信息检索期刊*(20), Springer, 第 338-362 页。
- Padmanabhan, B.、Fang, X.、Sahoo, N. 和 Burton-Jones, A. 2022. “信息系统研究中的机器学习”, *MIS Quarterly* (46:1), 第 iii-xix 页。

- Peffer, K.、Tuunanen, T. 和 Niehaves, B. 2018 年。“设计科学研究类型：适用设计科学研究范例和标准特刊简介”，《欧洲信息系统期刊》(27:2)，第 129-139 页。
- Peffer, K.、Tuunanen, T.、Rothenberger, MA 和 Chatterjee, S. 2007。“信息系统研究的设计科学研究方法”，《管理信息系统期刊》(24:3)，第 45-77 页。
- Piramuthu, S. 和 Doss, R. 2017。“基于传感器的多个 RFID 标签同时存在解决方案”，《决策支持系统》(95)，爱思唯尔，第 102-109 页。
- Prat, N.、Comyn-Wattiau, I. 和 Akoka, J. 2015。“信息系统构件评估方法分类”，《管理信息系统期刊》(32:3)，第 229-267 页。
- Ramakrishnan, M.、Gregor, S.、Shrestha, A. 和 Soar, J. 2023。“具有说明性实例的平台支持知识共享设计原则”，《信息系统协会期刊》(24:5)，第 1413-1438 页。
- Refsgaard, JC、Van der Sluijs, JP、Brown, J. 和 Van der Keur, P. 2006。“处理模型结构误差引起的不确定性的框架”，《水资源进展》(29:11)，爱思唯尔，第 1586-1597 页。
- Roberts, P. 和 Priest, H. 2006。“研究的可靠性和效度”，《护理标准》(20:44)，皇家护理学院出版公司 (RCN)，第 41-46 页。
- Rosemann, M. 和 Vessey, I. 2008。“提高信息系统研究与实践的相关性：适用性检查的作用”，《MIS Quarterly》(32:1)，第 1-22 页。
- Salmon, WC 1998。《因果关系与解释》，牛津大学出版社。
- Sedrakyan, G.、Poelmans, S. 和 Snoeck, M. 2017。“评估反馈包容的快速原型对新手建模者理解并行 UML 状态图语义的影响”，《信息和软件技术》(82)，爱思唯尔，第 159-172 页。
- Sein, M.、Henfridsson, O.、Purao, S.、Rossi, M. 和 Lindgren, R. 2011 年。“行动设计研究”，《MIS Quarterly》(35:1)，第 37 页。
- Shultz, KS, Riggs, ML 和 Kottke, JL 1998。“工业和人事心理学对效度概念不断演进的必要性：心理测量、法律和新兴问题”，《当代心理学》(17)，Springer，第 265-286 页。
- Taylor, CS 2013。《效度和确认》，《理解统计学》系列，纽约：美国牛津大学出版社。
- Thomas, MA、Li, Y. 和 Lee, AS 2022。“信息系统构件的一般化”，《信息系统研究》(33:4)，INFORMS，第 1452-1466 页。
- Tiefenbeck, V.、Goette, L.、Degen, K.、Tasic, V.、Fleisch, E.、Lalive, R. 和 Staake, T.，2016 年。“克服显著性偏差：实时反馈如何促进资源节约”，《管理科学》(64:3)，第 1458-1476 页。
- Tuunanen, T.、Winter, R. 和 Vom Brocke, J. 2024。“处理设计科学研究中的复杂性”，《MIS Quarterly》(48:2)，第 427-458 页。
- Umapathy, K.、Purao, S. 和 Barton, RR 2008。“设计企业集成解决方案：有效地”，《欧洲信息系统期刊》(17:5)，Springer，第 518-527 页。
- Valecha, R.、Sharman, R.、Rao, HR 和 Upadhyaya, S.，2013 年。“紧急响应系统的调度介导通信模型”，《ACM 管理信息系统汇刊 (TMIS)》(4:1)，ACM 纽约，纽约州，美国，第 1-25 页。
- Venable, J.、Pries-Heje, J. 和 Baskerville, R. 2016。“FEDS：设计科学研究评估框架”，《欧洲信息系统期刊》(25:1)，第 77-89 页。
- Venkatesh, V.、Morris, MG、Davis, GB 和 Davis, FD 2003。“用户对信息技术的接受度：走向统一观点”，《MIS Quarterly》，第 425-478 页。
- Weinhardt, C.、Kloker, S.、Hinz, O. 和 van der Aalst, WM 2020。“信息系统研究中

的公民科学”，*商业与信息系统工程*(62)，Springer，第 273-277 页。

Winter, S.、Berente, N.、Howison, J. 和 Butler, B. 2014 年。“超越组织‘容器’：21 世纪社会技术工作的概念化”，*信息与组织*(24:4)，第 250-269 页。

Zaitsev, A. 和 Mankinen, S. 2022。“设计金融教育应用促进发展：在柬埔寨农村应用行动设计研究”，*《欧洲信息系统期刊》*(31:1)，Taylor & Francis，第 91-111 页。

人物传记

Kai R. Larsen [0000-0002-8812-9866] 是科罗拉多大学博尔德分校利兹商学院的信息系统教授。他是信息科学系的兼职教员、阿格德大学的教授和盖洛普的研究顾问。Kai 最为人所知的是为 Edward Thorndike (1904) 的 Jingle Fallacy 提供了实用的解决方案，并为调查响应语义理论 (S TSR) 做出了贡献。他获得了 2019 年 INFORMS 设计科学奖、2020 年 Herbert A. Simon 设计科学奖和 2023 年 AIS 教育创新奖。

Roman Lukyanenko [0000-0001-8125-5918] 是弗吉尼亚大学麦金泰尔商学院的副教授。他的研究兴趣包括数据管理和研究方法（文献综述中的有效性和人工智能）。Roman 积极开发想法、工具和方法来改进数据管理和研究实践。这些解决方案获得了重大奖项，包括 INFORMS 设计科学奖、加拿大总督金奖、Hebert A. Simon 设计科学奖。Roman 的研究发表在《自然》、《MIS 季刊》、《信息系统研究》、《ACM 计算调查》上。他 2019 年关于众包数据质量的论文获得了 MIS 季刊最佳论文奖。

Roland M. Mueller [0000-0002-8706-7763] 是德国柏林经济与法律学院的信息系统教授。他获得了柏林自由大学的信息系统博士学位。他的研究兴趣包括用于文献综述的本体和大型语言模型、元理论分析、本体工程、设计科学方法以及 AI 项目中的用户驱动创新。他撰写了三本书和 100 多篇学术论文，其作品获得了多项荣誉，包括 Herbert A. Simon 设计科学奖。除了学术工作之外，

Roland 还作为多家初创公司的创始人和顾问委员会成员，为创新做出了贡献，并拥有机器学习专利。

Veda C. Storey [0000-0002-8753-1553] 是佐治亚州立大学 J. Mack Robinson 商学院的杰出大学教授、Tull 计算机信息系统教授和计算机科学教授。她的研究兴趣是数据管理、概念建模和设计科学。她特别感兴趣的是从数据管理的角度评估新技术对商业和社会的影响。Storey 博士是 AIS 高级学者学院和国际概念建模会议指导委员会的成员。她是 Peter P. Chen 奖的获得者、ER 研究员、AIS 研究员和 INFORMS 研究员。

Jeffrey Parsons [0000-0002-4819-2801] 是加拿大纽芬兰纪念大学工商管理学院的大学研究教授和信息系统教授。他的研究兴趣集中在如何更好地用数据来表达人类对世界的概念化。他在这方面和相关主题上的工作已在多个学科领域发表。Jeff 的研究得到了多项认可，包括年度 MISQ 论文 (2019)、AIS 资深学者论文奖 (2020) 和 INFORMS 设计科学奖 (2014)。他是信息系统协会会员、德累斯顿工业大学杰出研究员、Schoeller 资深研究员和 ER 研究员。他曾担任 MIS Quarterly 的资深编辑，目前是 Information Systems Research 的资深编辑。

Debra Vander Meer [0000-0002-5930-6667] 是佛罗里达国际大学 (FIU) 的信息系统教授。她目前的研究兴趣集中在将计算机科学和信息系统的概念应用于现实世界问题。她的作品发表在这些领域的顶级期刊和会议上。在加入 FIU 之前，她在业界工作了十年，包括在一家风险投资初创公司任职。她拥有乔治城大学的学士学位、亚利桑那大学的管理信息系统理学硕士学位以及佐治亚理工学院的计算机科学博士学位。

Dirk S. Hovorka [0000-0001-7049-5617] 是澳大利亚悉尼大学商业信息系统学科的教授。他获得了科罗拉多大学的信息系统博士学位，并拥有跨学科电信硕士学位和地质学硕士学位。他目前的研究探索了科学和社会实践如何通过理论、设计和未来概念化来产生可能的“世界”。他的研究重点是推测性地“认知”技术、社会和生物物理环境，认识到未来对经验性认知具有顽固的抵抗力，前所未有的社会技术变革使我们对过去的知识，对未来状态的指示性降低。Dirk 是 JAIS (研究观点) 的资深编辑、ISR 的编辑委员会成员，以及 HICSS “信息研究：参与未来”迷你轨道和 ECIS “未来作为探究场所”轨道的联合主席。

他是 2018 年 Beta Gamma Sigma 年度教授、韦恩·洛纳根杰出教学奖（2018 年）、院长教学奖（2022 年）的获奖者，并多次获得教学奖。Dirk 是 AIS 2011 最佳论文《二级设计：行为设计研究案例》的合著者。

Dr. Shu-Chen Chen **陈淑珍博士** 本文英翻中译编辑者，现任弗吉尼亚大学东亚语言文学文化系的中文副教授。拥有二十年的中文教学经验。2022 年荣获弗吉尼亚大学校友会首届优良教师奖。其博士论文专攻汉藏佛教净土法门研究。并在中文标准化考试评估领域具有专业造诣。

附录 A. 设计科学效度类型的总体使用情况

我们研究了设计科学效度框架，在 2004 年 4 月至 2017 年 12 月期间，在 AIS 资深学者八大期刊中发表的 199 篇设计科学论文中的使用程度，该框架转化为表 7 右栏中分类到我们框架中的现有 70 个效度定义。每篇论文都在句子级别进行检查，并与我们的效度类型列表进行比较，使用单词级别的正则表达式。第一作者阅读了所有结果句子，并排除了正则表达式产生过多假阳性的搜索结果。例如，诸如*准确性*、*完整性*、*精确度*和*召回率*在数据质量和机器学习研究中具有特定但多义的含义。其余术语产生了数万个句子，因此报告的结果是保守的估计。

我们确定了与框架中的 1 相对应的效度（表 7）在论文中至少使用一次的次数，按年份和效度类型进行分类。我们得出的结论是，设计科学论文使用的效度术语与其他学科不同，除了一些用于描述混淆矩阵中的指标的术语。与特征效度类型相关的效度名称很少使用。即使在使用率很高的功效效度类型类别中，讨论的大多数与效度相关的术语也源自机器学习中的混淆矩阵测量。*方法特征效度类型的使用是例外*，这是设计科学文献中使用率第二高的效度类型。

我们发现设计科学论文缺乏一致性，也缺乏对效度术语的实际使用，这意味着通过在明确的效度主张方面提高语言的一致性，有很大机会改善评估的沟通。一旦研究人员致力于共同的效度规范，沟通和报告的一致性就会得到改善。