

ОБОСНОВАННОСТЬ ЗНАНИЙ В НАУКЕ ДИЗАЙНА

Validity in Design Science

Кай Р. Ларсен

Школа бизнеса Лидса, Университет Колорадо, Боулдер, Колорадо 80309, США
{Kai.Larsen@colorado.edu}

Роман Лукьяненко

Школа коммерции Макинтайра, Университет Вирджинии, Шарлоттсвилль, Вирджиния 22908, США
{romanl@virginia.edu}

Роланд М. Мюллер

Кафедра бизнеса и экономики, Берлинская школа экономики и права, Берлин, Германия
{roland.mueller@hwr-berlin.de}

Веда С. Стори

Колледж бизнеса Дж. Мака Робинсона, Университет штата Джорджия, Атланта, Джорджия 30303, США {vstorey@gsu.edu}

Джеффри Парсонс

Факультет делового администрирования, Мемориальный университет Ньюфаундленда, Сент-Джонс, Канада {jeffreyp@mun.ca}

Дебра ВандерМеер

Колледж бизнеса, Международный университет Флориды, Майами, Флорида 33199, США
{vanderd@fiu.edu}

Дирк С. Ховорка

Школа бизнеса Сиднейского университета, Сиднейский университет, Сидней, Австралия
{dirk.hovorka@sydney.edu.au}

Исследователи должны быть уверенными, что утверждения о знаниях, полученных в результате их работы, являются обоснованными. Однако обоснованность - валидность - не является ни хорошо понимаемой, ни последовательно установленной в науке дизайна, которая включает в себя разработку и оценку артефактов (моделей, методов, экземпляров, и теорий). В результате сложно продемонстрировать и сообщить обоснованность утверждений о знаниях об артефактах. В этой статье определяется обоснованность в науке дизайна и выводится структура валидности науки дизайна и модель процесса для ее применения. Структура включает три основных типа утверждений и обоснованности — критериальный, причинный и контекстный, а также подтипы обоснованности. Структура помогает исследователям интегрировать соображения обоснованности в проекты, использующие науку дизайна, и вносит вклад в растущий объем исследований методологии науки дизайна. Она также предоставляет систематический способ формулирования и подтверждения утверждений о знаниях проектов науки дизайна. Для демонстрации самой структуры мы применили структуру к примерам из существующих исследований, а затем использовали ее для демонстрации обоснованности утверждений о знаниях самой структуры.

Ключевые слова: обоснованность знаний, валидность, наука дизайна, дизайн-наука, дизайнерская наука, научный дизайн, наука проектирования, научная методология проектирования, научная методология дизайна, структура валидности в науке дизайна, требование знаний, исследовательская валидность, критериальная валидность, причинная валидность, контекстная валидность, характеристическая валидность, валидность эффективности, внешняя валидность, экологическая валидность, оценка, валидация

Права на распространение: эта статья распространяется в открытом доступе <https://creativecommons.org/licenses/by-nc-nd/4.0/deed.en> по подписанному и выполненному соглашению с *MIS Quarterly*. Она распространяется в соответствии с правилами Национальных институтов здравоохранения США (NIH) для поддерживаемых исследований по гранту 3U24AG052175-08S1.

Цитировать: Ларсен, К., Лукьяненко, Р., Мюллер, Р., Стори, В. К., Парсонс, Дж., ВандерМеер, Д., and Ховорка, Д. 2025. «ОБОСНОВАННОСТЬ ЗНАНИЙ В НАУКЕ ДИЗАЙНА» *MIS Quarterly*, Перевод В. Лукьяненко, pp. 1–40. Английский оригинал: bit.ly/4inr9J9

Введение

Научные знания должны быть достоверными и надежными (Burton-Jones et al. 2021; Creswell and Miller 2000). Устоявшаяся традиция **валидности** (обоснованности знаний) способствует достоверной и надежной науке, поскольку она обеспечивает систематический способ оценки заявленного вклада в знание в научных исследованиях (Cook and Campbell 1979; Guba and Lincoln 1994). Каждая научная дисциплина имеет процедуры для оценки достоверности заявлений о знаниях. Формулирование этих процедур обеспечивает оценочную ясность, повышает эффективность исследований, способствует обмену передовым опытом, облегчает кумулятивную науку и способствует большему доверию общественности к науке.

В науке дизайна¹ — жанре исследований в области информационных систем,

компьютерных наук, инженерии, медицины, менеджмента, а также материаловедения и биомедицинских наук — исследователи выдвигают утверждения о знаниях относительно характеристик или производительности изобретенных ими артефактов. Исследователи предложили процессы для проведения науки дизайна, каждый из которых определяет оценку как существенный шаг (Gregor и Nevner 2013; Nevner et al. 2004; Peffers et al. 2007; Prat et al. 2015; Sein et al. 2011). Однако существующие методы и структуры не рассматривают обоснованность утверждений о знаниях как основной компонент оценки артефактов и не соглашаются с тем, что значит оценивать обоснованность утверждений о знаниях об артефактах.

Наука дизайна демонстрирует повторяющиеся типы утверждений о знаниях. Вот некоторые примеры: артефакт превосходит современное состояние дел (Padmanabhan et al. 2022); недавняя версия артефакта работает лучше, чем предыдущая версия (версии) (Sein et al. 2011); модель соответствует референтной системе реального мира (Gonzalez-Huerta et al. 2017);

¹ Design science (Design Science Research, Design Science Methodology, Scientific Design) имеет несколько возможных переводов на русский язык, в том числе **наука дизайна, дизайнерская наука, дизайн-наука, научный дизайн, наука проектирования, научная методология проектирования, научная методология дизайна.** Научный дизайн включает в себя **науку**

проектирования, поскольку научный дизайн также рассматривает создание готовых систем.

артефакт работает благодаря наличию определенных особенностей дизайна (Abbasi et al. 2012). Существует также набор общих подходов к оценке в науке дизайна (Prat et al. 2015; Venable et al. 2016), хотя они не затрагивают логику, которая связывает заявления с проверкой, процессом установления действительности определенного типа заявления о знаниях.

Хотя типы валидности используются в некоторых проектах по дизайну, они, как правило, адаптированы из других традиций. Например, внутренняя валидность и внешняя валидность адаптированы из психометрии (Baskerville et al. 2015; vom Brocke, Hevner, et al. 2020), тогда как метрики валидности, такие как точность и отзыв, взяты из поиска информации, статистики и других дисциплин (Larsen and Becker 2020). Исключением является валидность инстанцирования (Lukyanenko and Parsons 2020), которая была предложена как собственный тип валидности в дизайне.

Большинство валидаций в области науки дизайна не имеют названия, что затрудняет эффективное их использование и обмен передовым опытом с сообществом. Кроме того, некоторые процедуры оценки, такие как проверки применимости (Rosemann and Vessey 2008), отвечают на множественные (часто неуказанные) утверждения, что приводит к вопросам о *том, что точно* было подтверждено. Несмотря на потенциальные преимущества, не было никаких систематических попыток определить или исследовать валидность среди всех типов артефактов, рассматриваемых учеными дизайна (например, модели, методы, теории дизайна; см. Gregor and Hevner 2013) или предоставить процесс для проверки утверждений. Таким образом, основной вопрос о том, что составляет валидность в науке дизайна, остается без ответа.

Системный подход к валидности принесет пользу науке дизайна. Во-первых, он будет способствовать высококачественной научной работе. Решающее значение для качества науки имеет связь между утверждениями о

знаниях и доказательствами, которые подтверждают эти утверждения. Это может быть облегчено путем соблюдения процедур оценки в соответствии с согласованными структурами, которые поддерживают типы сделанных утверждений. Эти процедуры обычно требуют процедурной строгости и тщательной аргументации (например, доказательств причинно-следственной связи для внутренней валидности, часто предоставляемых посредством рандомизированных экспериментов). Установление обоснованности сделает связь между утверждениями и их подтверждением более явной и более поддающейся проверке. Эта прозрачность важна для формирования доверия и уверенности в утверждениях и сообщенных результатах (Burton-Jones et al. 2021).

Во-вторых, поскольку определенные типы валидности касаются определенных типов утверждений (например, внутренняя валидность касается утверждений о причинно-следственной связи), установление традиции валидности для науки дизайна приведет к тому, что утверждения о знаниях станут более явными. Это может помочь исследователям лучше оценить и сообщить о вкладе, который они вносят, и облегчить процесс валидации. Кроме того, развернутые утверждения о знаниях могут также использоваться другими исследователями, стремящимися расширить первоначальные вклады как в науке дизайна, так и за ее пределами, что позволит ученым-дизайнерам вносить более весомый вклад в кумулятивную традицию исследований в области проектирования, использования и воздействия артефактов.

В-третьих, всесторонняя концептуализация валидности увеличит влияние науки дизайна на реальный мир. Благодаря повышению чувствительности исследователей к природе утверждений о знаниях и типам валидности, артефакты, которые не выполняют своего предназначения при воздействии неожиданных обстоятельств реального мира, менее вероятны. Например, в машинном обучении исследователям важно знать, какие

особенности артефакта дали улучшенные результаты. Тем не менее, они узко сосредоточены на производительности предлагаемого артефакта. Более инклюзивная концепция валидности должна учитывать применимость артефакта для различных условий реального мира, в том числе для более широких задач и различных групп пользователей (Ethayarajh and Jurafsky 2020).

В-четвертых, установление традиции валидности будет способствовать публикации исследований по дизайну. Как отражение общественных норм и общих концепций, традиция валидности повышает производительность исследований за счет повышения согласованности валидации (Chan 2014) и оптимизации практик валидации. Обратите внимание, что плодотворные области эконометрики и психометрии имеют устоявшиеся подходы к валидности или валидации (Taylor 2013), как и качественные исследования (Creswell и Miller 2000) и области компьютерной науки и программной инженерии, такие как моделирование (например, Pääkkönen et al. 2017) и разрешимость проблем (например, Fan et al. 2018). В дисциплинах более высокий уровень согласия по основным дисциплинарным нормам (включая валидацию) коррелирует с повышенной производительностью исследований и качеством публикаций (Gumpert 2007). Формализация общих дисциплинарных концептуализаций имеет жизненно важное значение для ускорения научного прогресса (Национальные академии наук 2022 г.).

Наконец, установление природы валидности в науке дизайна позволит позиционировать ее по отношению к устоявшимся традициям (например, в поведенческих науках) и лучше донести до других дисциплин вклады науки дизайна. Общие шаблоны валидации получают систематическое название, тем самым помогая ученым и посторонним в области науки дизайна ссылаться на эти подходы к оценке. Это, в свою очередь, способствует установлению идентичности науки дизайна.

Это исследование вносит несколько вкладов. Сначала мы изучаем, как концепции валидности использовались в науке дизайна. Затем мы рассматриваем общую природу валидности, чтобы определить установленные основы, которые могут принести пользу ученым дизайна. Мы делаем это, развивая понимание логики и взаимосвязей между утверждениями, типами артефактов, типами валидности и контекстом оценки. Затем эти основы позволяют нам определять типы валидности и использовать их соответствующим образом для оценки утверждений о научных знаниях в области дизайна.

Далее, мы разрабатываем структуру валидности в науке дизайна Design Science Validity Framework, чтобы обеспечить систематический способ формулирования и подтверждения заявлений о знаниях в проектах научного дизайна. Структура включает три типа требований к знаниям и валидности — *критериальные*, *каузальные* и *контекстные* — а также подтипы валидности. Эта структура соответствует цели науки дизайна разрабатывать инновационные артефакты как решения социальных проблем, одновременно создавая научные знания, которые практики могут повторно использовать в различных контекстах. Структура является инклюзивной в своем охвате артефактов науки дизайна, включая реализованные системы (например, инструменты и развернутые экземпляры), абстрактные вклады, которые включают чертежи для разработки систем (например, концептуальные модели или методы машинного обучения), и теоретические знания дизайна (например, теории дизайна, которые предписывают дизайн и действие для достижения определенных целей, см. Gregor and Jones, 2007).

Наконец, мы оценили структуру валидности в науке дизайна с помощью серии исследований, рекурсивно информированных самой структурой. С помощью двух проверок применимости, обширного анализа исследований валидности и всестороннего анализа оценки науки дизайна мы

демонстрируем полезность нашей структуры для исследователей, ее способность охватывать существующие нормы валидации и ее экономичность. Мы завершаем статью общим обсуждением природы и важности валидности науки дизайна, рекомендуем, как можно использовать структуру, и предлагаем будущие возможности для исследований.

достоверным и надежным практическим научным знаниям. Эти усилия были сосредоточены на обеспечении качества в разработке артефактов и оценке полезности и качества артефактов и знаний о дизайне. (См. Таблицу 1 для конкретных рассматриваемых тем). В то время как строгости в разработке артефактов уделялось много внимания, строгость в оценке полезности и качества артефактов и знаний о дизайне отстает, хотя это и является серьезной проблемой.

МЕТОДОЛОГИЧЕСКИЕ РАЗВИТИЯ В НАУКЕ ДИЗАЙНА

Наука дизайна разработала методологические основы для содействия

Таблица 1. Методологические основы в дизайн-науке	
Вклад	Ссылка
Интеграция разработки систем со строгой теорией и эмпирической оценкой	Нунамейкер и др. (1991) ; Марч и Смит (1995)
Подходы к оценке артефактов и теорий	Хевнер и др. (2004) ; Венейбл и др. (2016) ; Прат и др. (2015) ; Грегор и Джонс (2007) ; Туунанен и др. (2024)
Методы создания и распространения науки дизайна	Пефферс и др. (2007) ; Грегор и Хевнер (2013) ; Йоханнессон и Перйонс (2014) ; Баскервиль и др. (2015) ; Авиджи и Винтер (2019) ; Иивари (2015) ; Грегор и др. (2020) ; Туунанен и др. (2024)
Сопоставление основных теорий с артефактами	Арази и др. (2010) ; Грегор и Джонс (2007) ; Кюхлер и Вайшнави (2012) ; фон Брокке, Винтер и др. (2020)
Прозрачность	фон Брокке и др. (2021) ; Бертон-Джонс и др. (2021) ; Лукьяненко и Парсонс (2020) ; Хевнер и др. (2024)
Соответствие практике	Сейн и др. (2011) ; Лукьяненко и Парсонс (2020) ; Хевнер и др. (2024)
Теоретизация дизайна	Мандвивалла (2015) ; Грегор и Джонс (2007) ; Грегори и Мюнтерманн (2014) ; Лукьяненко и Парсонс (2020) ; Грегор и др. (2020)

Значительным шагом к установлению качества в науке дизайна является понимание того, какие валидации уже существуют, как они предоставляют дополнительные знания и как они пересекаются. Ни одна из предыдущих работ не пыталась определить качество оценки науки дизайна путем установления возможных типов валидности. Для этого требуется понимание типов

утверждений о знаниях, артефактов и подходов к валидации.

Мы рассмотрели известные работы по основам дизайна, которые обобщены в Таблице 2. Большинство основ в Таблице 2 в первую очередь касаются строгости разработки артефактов. Например, Сейн и др. (2011) сосредоточились на исследовании дизайна действий (ADR), где оценки

происходят в реальном, аутентичном контексте. Венейбл и др. (2016) различали формирующие и суммирующие оценки, уделяя особое внимание тому, когда проводятся оценки. Грегор и Хевнер (2013) рассмотрели способы позиционирования и представления науки дизайна и то, происходят ли оценки внутри или вне контекста разработки. Прат (2015) организовали подходы к оценке в таксономию пяти измерений различных типов оценки. Баскервиль и др. (2015) разработали основу для выделения различных типов производства знаний, которые могут происходить в ходе проекта по дизайну. Туунанен и др. (2024) предоставили подход к управлению сложными, многоэтапными проектами, предположив, что эти этапы («эшелоны») производят различные артефакты, которые могут быть оценены на каждом этапе.

Примечательно, что проблемы валидности со временем выросли от незначительного или нулевого рассмотрения в ранних фреймворках (например, Hevner et al. 2004; Peffers et al. 2007) до активного рассмотрения в недавних методологических исследованиях (например, Baskerville et al. 2015; Tuunanen et al. 2024).

Популярные фреймворки в Таблице 2 представляют собой устоявшееся методологическое руководство в области

науки дизайна. Они обеспечивают ясность для понимания общего подхода к проведению исследований и оценке артефактов, но не предлагают руководства по валидационным мероприятиям, включая то, как можно предоставить доказательства для обоснования заявлений о знаниях.

Таблица 2. Фреймворки в науке дизайна и роль валидности

Статья	Обзор фреймворка	Валидность	Оценка
Хевнер и др. (2004)	Наука дизайна заключается в создании и оценке артефактов, а также в их соединении с окружающей средой и базой знаний дизайна.	Минимальный акцент на обоснованности. Ссылаются на другие работы о том, как накопление доказательств в конечном итоге установит обоснованность более широких утверждений о дизайне.	Предлагает два фундаментальных вопроса, которые требуют доказательств от оценок: 1) какую полезность обеспечивает артефакт? и 2) как эта полезность демонстрируется? Предлагает, что вклад возникает из полезности, продемонстрированной посредством оценки.
Пэфферс и др. (2007)	Шестиэтапный итеративный процесс представления и оценки научных разработок в области дизайна.	Мало внимания обоснованности. Предполагает, что наука дизайна должна была использовать «аргументы ad hoc» (стр. 50).	Исследователи наблюдают и измеряют, насколько хорошо артефакт поддерживает решение. Включает сравнение целей решения с фактическими наблюдаемыми результатами использования артефакта (например, через время отклика, произведенные элементы, опросы удовлетворенности пользователей и моделирование).
Сейн и др. (2011)	Четырехэтапная модель исследования дизайна действий.	Не содержит упоминаний о валидации.	Приводит доводы в пользу принципа «подлинной и одновременной оценки», чтобы подчеркнуть ключевую характеристику

			исследования дизайна действий.
Грегор и Хевнер (2013)	Схема публикации в области науки о дизайне, включая оценку.	Оценка артефакта с точки зрения критериев, которые могут включать в себя валидность, полезность, качество и эффективность.	Различает оценку внутри и вне среды разработки. Критерии полезности, связанные с тем, переносятся ли производительность за пределы среды разработки.
Прат и др. (2015)	Таксономия методов оценки с пятью измерениями: цель, среда, структура, деятельность и оценка.	Валидность является частью целевого измерения эффективности и результативности. Валидность достигается, если артефакт работает правильно (достигает цели).	Сосредоточение на «что» (системы артефактов) и «как» (используемые методы) в оценке. Оценка включает в себя «относительность» превосходства артефакта над другими решениями. Определяет типичные методы оценки и вторичные оценки.
Баскервиль и др. (2015)	Жанры исследования. Структура: дизайн против науки и номотетическое против идиографического.	Предоставляет 18 критериев качества. Упоминает внутреннюю и внешнюю валидность.	Предлагает размышления о жанрах исследования, а не об оценке как таковой.
Венейбл и др. (2016)	Оценки в области науки дизайна существуют в двух измерениях: искусственное против натуралистического и формирующее против суммирующего.	Валидность рассматривается как результат силы оценки и того, насколько хорошо артефакт достигает своей цели.	Основная цель оценки — определить, насколько хорошо артефакт достигает предполагаемой полезности.

Некоторые работы в области науки дизайна признали важность установления валидности (vom Brocke, Winter и др. 2020) и предположили, что существующие типы поведенческой валидности (например, внутренняя, экологическая) могут применяться к науке о дизайне (Baskerville и др. 2015). В науке дизайна, использующей методы на основе искусственного интеллекта, такие как машинное обучение и обработка естественного языка, принято сообщать о

точности, полноте и F_1 -балле, а также других метриках матрицы путаницы (например, Abbasi и Chen 2008; Li и др. 2020). Это не типы валидности, но они связаны с валидностью, поскольку они предоставляют количественные меры для установления валидности утверждений. Они также не применимы ко всем артефактам науки о дизайне (например, теориям дизайна). Чтобы установить валидность в науке дизайна, мы сначала рассматриваем валидность в науке в

целом, которую затем синтезируем с конкретными проблемами науки дизайна.²

ДОСТОВЕРНОСТЬ ИССЛЕДОВАНИЯ

Основная идея о том, что утверждения о знаниях должны быть подтверждены, восходит к античности (Carter 2019). Термин «валидность» (*validity*) возник в количественных социальных науках (например, Cronbach and Meehl 1955) и позже стал более широко использоваться, поскольку он применялся к развивающимся междисциплинарным убеждениям о создании соответствующих оценок (например, в которых оцениваются предыдущие знания и согласовываются критерии приемлемости) для установления обоснованности утверждений о знаниях.

Чтобы понять валидность в науке дизайна, мы изучили около 7500 источников из литературы по валидности исследований в области информационных систем (включая науку дизайна) и других областей исследований, таких как компьютерные науки, социальные науки, инженерия, медицина, юриспруденция и гуманитарные науки. Эта работа дала 2418 кандидатов на валидность, что составляет самую большую работу по пересмотру валидности более чем на порядок. Мы сообщим об этих усилиях позже, когда проверим нашу структуру. Здесь мы кратко рассмотрим, как валидность используется в науке, поскольку это информирует наши представления о валидности науки дизайна.

Первоначально валидность узко понималась как «близость того, что, как мы считаем, мы измеряем, к тому, что мы намеревались измерить» (Roberts and Priest 2006, p. 41). Эта точка зрения фокусировала валидность на артефактах измерения, особенно в психометрии, связанной с тестами,

инструментами или анкетами, которые назначались людям. Со временем количественные исследователи разработали много различных видов валидности (например, внутреннюю, экологическую, дискриминантную, внешнюю), которые теперь широко используются в поведенческих исследованиях информационных систем (Boudreau et al. 2001). Использование этих валидностей было признано «профессиональной ответственностью» (Shultz et al. 1998, p. 266).

Качественные и интерпретативистские исследователи отстаивали уникальные проблемы валидности в своей работе, подчеркивая важность контекста, обстановки и участников, в дополнение к роли исследователей в создании естественного, заслуживающего доверия, подтверждаемого и надежного отчета о процессе исследования (Lincoln and Guba 1985; Onwuegbuzie and Leech 2007). Валидность в этой традиции считается продуктом социального консенсуса, где то, что является валидным, основано на «общепринятой приемлемости» (Moules et al. 2015, p. 172). Размышляя об этих усилиях, Кресвелл и Миллер (2000, стр. 124) отметили «общее согласие» относительно того, что валидность является основополагающим фактором для принятия исследований — точка зрения, широко распространенная во многих дисциплинах (Houningen-Huene 2013), включая информационные системы (Burton-Jones et al. 2021).

В компьютерной науке, информатике, дисциплине, тесно связанной с проектированием в информационных системах, исследователи в таких подобластях, как машинное обучение и искусственный интеллект, приняли общую структуру задач, сосредоточившись на общих контрольных показателях для проверки новых артефактов относительно современного уровня техники, отраженного в общих таблицах лидеров (Matadamas-Hernández et al. 2012). Эта структура стала катализатором крупных достижений в

² В Приложении А оценивается текущее состояние практики валидации.

области компьютерного зрения и обработки естественного языка. Например, исследователи машинного обучения тщательно разработали подходы, такие как анализ причинно-следственной связи и исследования абляции, для оценки разработанных артефактов (например, Chowdary и Kanhangad 2022). Они используют эти подходы в рамках общей структуры задач для проверки утверждений о знаниях и вносят вклад в общие знания о задачах. Тем не менее, несмотря на впечатляющий прогресс, общая структура задач могла ограничить практическое применение полученных моделей и ухудшить критерии оценки, такие как «компактность, справедливость и энергоэффективность» (Ethayarajh и Jurafsky 2020, стр. 1) .

Каждый тип исследования, а иногда и каждая область (например, машинное обучение) разработали отдельные традиции валидности. Так же, как различаются вопросы качественной и количественной валидности, можно ожидать, что валидность в науке дизайна будет иметь свою собственную направленность. Помимо простого заимствования типов валидности из психометрических, компьютерных или качественных исследований, есть возможность обратиться к уникальной природе науки дизайна, в которой исследователи разрабатывают артефакты для получения желаемых результатов и подтверждения утверждений о знаниях об этих артефактах. Тем не менее, существующие исследования валидности в других дисциплинах представляют ценность для валидности в науке дизайна, поскольку они имеют дело с общими концепциями и идеями валидности. Наш обзор литературы по валидности привел к следующим выводам, которые информируют о валидности в науке дизайна.

Во-первых, исследователи согласны с тем, что систематически оцениваемые и подтверждаемые научные исследования необходимы (Cohen et al. 2013). Даже исследователи, которые приходят к выводу,

что название «валидность» не подходит для их исследований, признают преимущества выделения и распространения успешных и проверенных моделей валидации (например, Creswell and Miller 2000; Maxwell 1992).

Во-вторых, все науки содержат утверждения о знаниях (Collier-Reed и Ingerman 2013). Исследование может иметь много утверждений о знаниях, некоторые из которых являются фокусом исследования (первичные утверждения), а другие взяты из предыдущих исследований (вторичные утверждения). Часто только первое подлежит проверке в исследовании. Таким образом, в контексте валидности **утверждение о знаниях** является утверждением об интересующих явлениях, которое фиксирует оригинальный вклад исследования.

В-третьих, при проверке утверждения о знаниях объект утверждения обычно сравнивается с критерием — абстрактным или материальным объектом, свойства или поведение которого можно сравнить с предлагаемой идеей или объектом для оценки качества последнего. Например, в информатике функциональная валидность достигается, когда «модель имитирует поведение ввода-вывода реальной системы с некоторым приемлемым уровнем точности» (Murphy-Smith, 2015, стр. 30). Здесь реальная система является критерием. Критерии существуют в различных формах, включая артефакты, существующие естественные объекты (например, человек) и ментальные идеи. Независимо от исследовательского утверждения, всегда существует критерий, с которого можно проверить утверждение. Даже для радикальных инноваций и изобретений существуют критерии, с которыми их можно сравнить.³

³ Рассмотрим великие изобретения человеческой истории, включая огонь, колесо, динамит, гвоздь и печатный станок. У всех были ручные или менее инновационные артефакты, по которым их можно было оценить. Хотя гениальность многих революционных изобретений заключается в том, в какой степени они отходят от существующих артефактов или способов выполнения задач

В-четвертых, для подтверждения достоверности знания исследователи проводят одну или несколько оценок. Процедура оценки (валидации) — это набор задач, предпринимаемых для предоставления доказательств обоснованности утверждения о знаниях. Обычно процедура включает сравнение по некоторым параметрам с критерием, который может быть материальным или абстрактным. Эти процедуры обычно устанавливаются на основе консенсуса в рамках дисциплины (Тейлор 2013).

В-пятых, валидность — это вопрос степени, поскольку уместность и качество критерия и подход, используемый для его валидации, могут различаться. Таким образом, особенно сильные сравнения являются консервативными — выполняются с учетом лучших знаний, *современного* артефакта или способа выполнения чего-либо, известного как эталон. Как количественные, так и качественные исследователи (например, Moules et al. 2015; Taylor 2013) утверждают, что достижение идеальной валидности невозможно. Каждое исследовательское сообщество определяет нормы того, что составляет достаточный результат сравнения, чтобы соответствующее утверждение о знании могло быть принято как научно обоснованное.

Наконец, чтобы систематизировать согласованные и широко используемые шаблоны процедур оценки и норм для принятия доказательств для определенных типов утверждений о знаниях, они маркируются и организуются в типы валидности. Затем утверждение определенной валидности (например, конвергентной валидности) может использоваться в качестве сокращения для предположения, что соответствующее

(референтной сущности), когда ни один существующий референтный артефакт не может быть идентифицирован как критерий или к нему невозможно получить доступ (например, он является собственностью), тогда можно рассмотреть другие соответствующие артефакты или процессы.

утверждение (что две меры, по сути, связаны) может быть принято как доказанное на основе существующих научных норм. Например, Кук и Кэмпбелл (1979) исследовали угрозы и представили подходы для обеспечения внутренней валидности в социальных науках, в то время как Линкольн и Губа (1985) предложили шаги для демонстрации надежности и подлинности утверждений в качественных исследованиях. Эти шаблоны валидности предлагают конкретные подходы для определения критериев, их сравнения и представления результатов. Это упрощает ссылку и обмен нормами валидации, их последовательное использование и улучшение путем выявления связей, совпадений и пробелов между типами валидности, тем самым способствуя кумулятивной научной традиции.

СТРУКТУРА ВАЛИДНОСТИ НАУЧНОГО ДИЗАЙНА

Учитывая основы валидности, мы теперь разрабатываем Design Science Validity Framework.⁴ Мы начинаем с утверждений о знаниях о создаваемом артефакте. Объяснение этих утверждений может помочь ученым в обмене и повторном использовании полезных процедур валидации.

Утверждения о научных знаниях в области дизайна

Природа утверждений о дизайне как науке коренится в науке дизайна как утилитарном научном исследовании (Hevner et al. 2004). Ключевая цель науки дизайна — создавать артефакты, которые решают реальные проблемы, и генерировать знания о дизайне. Например, Хевнер и др. (2004, стр. 77) определили науку дизайна как исследование, которое «создает и оценивает ИТ-артефакты,

⁴ Далее мы будем называть структуру валидности научного дизайна просто «структурой валидности», если только мы намеренно не будем называть ее полным названием.

предназначенные для решения выявленных организационных проблем». С тех пор понимание науки дизайна вышло за рамки чисто организационной направленности, поскольку исследователи все чаще стремятся решать более широкие общественные и индивидуальные проблемы (Weinhardt et al. 2020; Winter et al. 2014). Кроме того, многие артефакты являются социотехническими, связывая программное обеспечение, оборудование и процессы с отдельными лицами, группами и организациями (Thomas et al. 2022). Таким образом, наука дизайна охватывает широкий спектр социальных проблем и соответствующих инновационных артефактов. Что отличает науку дизайна от практики, так это цель разработки знаний, связанных со строительством артефактов (Gregor and Jones 2007; Peffers et al. 2018). Поэтому мы определяем науку дизайна как исследование, которое разрабатывает новые артефакты и соответствующие знания в области дизайна для решения индивидуальных, организационных и общественных проблем.

Многие типы артефактов являются вкладами в науку дизайна. Артефакты — это модели, методы, экземпляры и теории дизайна и компоненты теории (например, конструкции, принципы дизайна) (Gregor и Hevner 2013). Артефакты — это сложные объекты с несколькими взаимосвязанными частями или компонентами.⁵ Их также обычно считают компонентами более крупных систем, таких как организации или более широкие технические системы. Учитывая это разнообразие, мы определяем **артефакт науки дизайна** как абстрактную или конкретную сущность, созданную как центральный вклад проекта науки дизайна для достижения желаемых результатов на индивидуальном, организационном или общественном уровне. Базовые принципы, согласно которым наука дизайна

предоставляет решения проблем и генерирует предписывающие знания о дизайне, являются центральными для утверждений о знаниях в области дизайна. В частности, для оценки вклада артефакта, артефакт должен быть в некотором роде лучше (например, быстрее, эффективнее, дешевле) по сравнению с существующими решениями (Padmanabhan et al. 2022), а подход, используемый для создания артефакта, должен быть указан (Hevner et al. 2004). Чтобы функционировать как наука и позволить другим извлекать выгоду из накопленных знаний о дизайне, исследователям необходимо передавать знания о том, как был создан артефакт, например, путем обмена кодом в репозитории или описания того, как артефакт функционирует на концептуальном уровне (Burton-Jones et al. 2021; Hevner et al. 2024).

Кроме того, конкретные причинно-следственные механизмы, связывающие выбор дизайна (части артефакта) с желаемыми результатами, могут быть предоставлены для углубления знаний о дизайне, возникающих в результате разработки артефакта (Gregor and Jones 2007; Peffers et al. 2018). Наконец, поскольку практики вряд ли будут реализовывать свои решения в контекстах, идентичных исходному исследовательскому контексту, представленному в статье, важно указать, когда и при каких условиях и границах артефакт, как ожидается, достигнет своих результатов (Hevner et al. 2024). Ключевой задачей для ученых-дизайнеров является поддержка других исследователей и практиков в повторном использовании артефактов и знаний о дизайне в новых условиях (например, Iivari et al. 2021).

Деятельность по проведению науки дизайна порождает определенные типы утверждений о знаниях.

⁵ Каждый тип артефакта также имеет много подтипов. Например, подтипы *модели* включают фреймворки, таксономии, онтологии, симуляции и математические модели.

Утверждение о знании в науке дизайна — это предложение об артефакте, которое утверждает свой вклад в науку и общество посредством определенной формы или

функции. Например, Тифенбек и др. (2016) разработали счетчик для душа и заявили, что конкретное указание потребляемой воды и энергии (форма артефакта) снижает потребление энергии (функция артефакта) по сравнению с душами, не имеющими таких показателей.⁶ В соответствии с целями, утверждения о знании в науке дизайна предполагают, что артефакт, созданный с помощью определенного подхода, приносит желаемый результат, что результат вызван определенными свойствами дизайна артефакта или что результат, как ожидается, будет сохраняться в определенном контексте или в различных контекстах. Мы обозначаем эти типы утверждений о знании как *критериальные*, *причинные* и *контекстные* соответственно и подробно останавливаемся на них ниже.

Критериальные утверждения: В каждой рассмотренной нами научной работе по дизайну авторы делают одно или несколько утверждений об *ожидаемых результатах* артефакта, таких как его преимущества или полезность для решения проблемы. Мы определяем *критериальное утверждение* как утверждение о знаниях о полезности артефакта. *Всегда* есть существующий или альтернативный способ сделать что-либо, поэтому критериальные утверждения напрямую или косвенно представляют сравнение с такими существующими сущностями или процессами (также известными как состояние дел или критерий) (Padmanabhan et al. 2022). Например, авторы могут утверждать, что разработали более эффективную поисковую систему (Li et al. 2020) или новую теорию дизайна

настраиваемых технологий (Germonprez et al. 2007). Такие утверждения предполагают, что, соответственно, обычные поисковые системы могли бы работать лучше, извлекая теоретические конструкции из статей, или что типичные теории дизайна не рассматривают настраиваемые технологии. Мы называем эти утверждения критериальными утверждениями, даже если само утверждение может явно не указывать критерий.

Критериальные утверждения особенно сильны, когда артефакт сравнения является общепринятым современным уровнем техники (Padmanabhan et al. 2022). Например, Larsen и Bong (2016) сравнили производительность своего алгоритма поиска с поисковой системой EBSCO Host, тогда как Li et al. (2020) сделали, возможно, более сильное утверждение, сравнив свой алгоритм поиска как с EBSCO Host, так и с Google Scholar — общепризнанным современным уровнем техники для академического поиска. В нашем исследовании статей по дизайну (о котором будет сообщено позже) мы обнаружили, что *во всех статьях* были критериальные утверждения, которые неявно или явно указывали на результаты, которые были лучше, чем существующие артефакты или процессы.

Причинные утверждения: В долгосрочной перспективе критериальное утверждение недостаточно для обеспечения научного прогресса. После того, как критериальные утверждения были установлены, исследователи могут (в рамках исследования, устанавливающего критериальное утверждение, или в последующих исследованиях) захотеть углубить знания о дизайне, обосновав утверждения о том, какие конструктивные особенности артефакта приводят к их заявленным результатам. Мы называем такие утверждения *причинные утверждения*, учитывая фундаментальный интерес науки к причинам и следствиям (Salmon 1998). Причинное утверждение — это утверждение о знании того, в какой степени конкретные части артефакта вызывают определенный результат.

⁶ В контексте теорий дизайна Грегор и Джонс (2006, стр. 327) выдвинули связанное понятие «проверяемых утверждений или гипотез», объяснив, что эти утверждения о теориях дизайна проверяются «через реализацию, путем построения системы или реализации метода, или, возможно, в редких случаях, посредством дедуктивной логики». Утверждение о знании дизайна — более широкое понятие: как мы показываем в статье, утверждения о знании применяются к любому артефакту, и для их подтверждения не требуется реализация их компонентов, даже в случае теории дизайна.

В то время как предоставление кода или описаний артефакта помогает поддерживать ценность артефакта и делает его более доступным для репликации и адаптации, продвижение причинных утверждений может помочь построить лучшие артефакты и разработать новые теории технологий. Например, Аббаси и Чен (2008) выдвинули несколько причинных утверждений, когда заявили, что наборы функций CyberGate лучше представляли типы информации, чем базовые наборы функций, которые обычно используются в системах (обратите внимание на наличие критериального утверждения). Они провели отдельные оценки для каждого из расширенных наборов функций, а именно темы, мнения, стиля, жанра и информации о взаимодействии, чтобы установить причинно-следственное влияние каждого из них на производительность артефакта (критерий исходит из производительности артефакта).

Причинные утверждения не ограничиваются машинным обучением или даже экземплярами артефактов. Так, позже мы покажем как мы оцениваем причинные утверждения о ценности частей нашей собственной структуры валидности науки дизайна. Фактически, учитывая, что большинство типов артефактов науки дизайна не функционирующие системы (а, например, фреймворки или концептуальные модели), качественные методы играют важную роль в оценке артефактов. Эти методы, в свою очередь, могут быть основаны на оправдательных знаниях (например, теории), которые предполагают причинные механизмы, связывающие особенности артефактов с целевыми результатами (Gregor and Jones 2007).

Не всегда возможно узнать, как или почему что-то работает, особенно когда речь идет о сложных технологиях, инновационных артефактах и «изобретениях» (Gregor and Nevner, 2013). Следовательно, не все статьи делают причинные утверждения; в нашем обзоре статей по дизайну в 50% были сделаны причинно-следственные заявления.

Контекстные утверждения: Артефакты всегда создаются в каком-то контексте для решения заданной проблемы. Однако роль контекста не всегда учитывается в науке дизайна. Чтобы помочь практикам в реализации артефактов, *контекстные утверждения* объясняют ситуации или условия, в которых, как ожидается, будут иметь место предлагаемые результаты артефакта. Соответственно, критериальные и причинные утверждения обычно считаются более сильными, когда оцениваются в нескольких контекстах или контекстах, которые больше похожи на предполагаемый контекст использования артефакта.

Не все научные статьи по дизайну делают заявления о контексте. Исключением является исследование дизайна действий (action design research), где обстановка, в которой реализуется артефакт, реальная организация (Sein et al. 2011), является контекстом, в котором любые заявления об артефакте должны применяться. Заявление о контексте может быть расширено (и, следовательно, сфера действия критерия или причинных заявлений усилена) с предложениями о том, что артефакт работает за пределами своей первоначальной обстановки, например, в схожих или даже отличных условиях. Кроме того, заявление о контексте может помочь углубить обобщение результатов. Например, Лукьяненко и др. (2019) прямо заявили, что их метод сбора данных на основе экземпляров особенно эффективен для задач в крупномасштабных проектах гражданской науки, но также заявили, что он не должен давать преимуществ для сбора в краудсорсинге микротаск. В нашей выборке научных статей по дизайну 14% статей содержали утверждения о контексте. Это число могло быть занижено авторами, которые явно не делали таких утверждений, даже когда было вероятно, что такие утверждения могли быть подкреплены предоставленными доказательствами. Если это так, то это дает возможность исследователям рассмотреть возможность продвижения утверждений о контексте (чтобы другие, читающие их работу, могли лучше оценить другие

контексты, в которых разработанные ими решения также могли быть полезны).

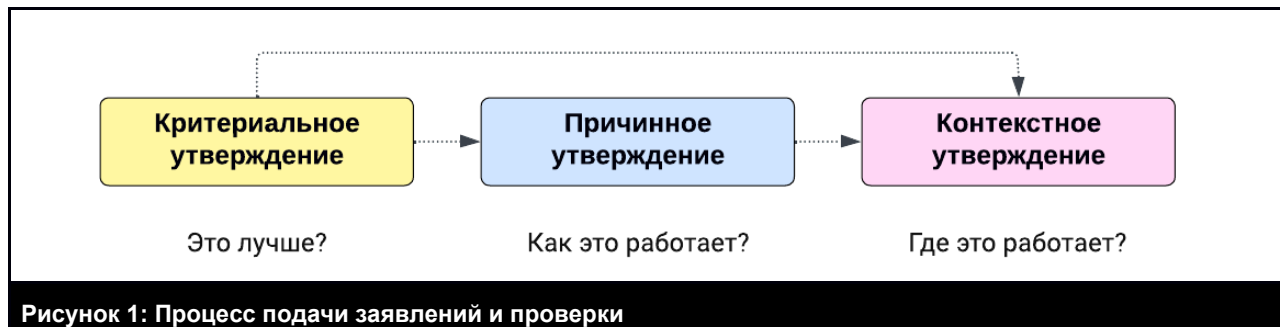


Рисунок 1 суммирует процесс проверки на основе важности и распространенности утверждений о знаниях. Рисунок показывает, что каждая статья делает по крайней мере одно критериальное утверждение. Этого может быть достаточно для новых артефактов. Как предполагают Грегор и Хевнер (2013, стр. 346), новые артефакты являются «интересными приложениями, где существует мало текущего понимания контекста проблемы и где нет эффективных артефактов в качестве решений». Затем, в зависимости от новизны артефакта, статья может дополнительно продвигать причинные и/или контекстные утверждения, чтобы улучшить понимание того, почему и где артефакт работает. Причинные утверждения требуют одновременного или предварительного установления критериального утверждения; контекстные утверждения требуют либо установления критериального утверждения, либо причинного утверждения.

Три типа заявлений о знаниях могут быть сделаны в отношении всех типов артефактов. Как обсуждалось, абстрактные артефакты (например, теория дизайна, концептуальные модели) могут быть инстанцированы и (через инстанциацию) может быть показано, что они оказывают измеримое влияние на мир. Кроме того, качественные методы, такие как контрфактуальный анализ и интервью с заинтересованными сторонами, могут использоваться для оценки критериальных, причинных и контекстных утверждений для абстрактных артефактов.

Типы утверждений определяются желаемым вкладом в знания, а не типом артефакта.

Артефакты и сравнительные критерии науки о дизайне

Мы определяем **валидность** как степень, в которой утверждения о знаниях о создаваемом артефакте подтверждаются в науке дизайна доказательствами. В соответствии с тремя типами утверждений о знаниях существуют три общих типа обоснованности науки дизайна: критериальная, причинная и контекстная. Поддерживая свои соответствующие утверждения, эти типы обоснованности продвигают практику науки дизайна тремя важными способами. Во-первых, *критериальная валидность* используется для поддержки утверждений о результатах проектирования и использования артефакта, направленных на достижение желаемых изменений. Во-вторых, *причинная валидность* используется для поддержки утверждений о вкладе конкретного дизайна (особенности артефакта, части) в полезность артефакта. В-третьих, *контекстная валидность* касается степени, в которой утверждения о знаниях верны для предполагаемого контекста или дополнительных контекстов. В рамках этих широких категорий мы предлагаем более конкретные типы обоснованности, основанные, в первую очередь, на типе сравнения и, во-вторых, на характере сравнения между артефактом и референтом.

Некоторые артефакты являются *материальными*, а другие *абстрактными* (Gregor and Hevner 2013, стр. 341). Например, душевой счетчик Тифенбека и др. (2016) является материальным, тогда как теория, представляющая собой систему концепций и предложений (Gregor and Jones 2007), является абстрактным артефактом.

Абстрактные артефакты могут использоваться при разработке материальных артефактов, таких как счетчики душа. Часто существуют *два отдельных, но связанных артефакта* — абстрактный (например, проектные предложения) и материальный артефакт, созданный на ее основе (например, счетчик душа). Артефакты являются компонентами более крупных систем, таких как *социотехнические системы* людей и артефактов (Chatterjee et al. 2021; Winter et al. 2014). Для многих проектов важно учитывать более широкие системы, в которые встроены артефакты, поскольку заявления о знаниях могут быть сделаны в отношении воздействий на уровне систем. Если исследователь стремится спроектировать саму социотехническую систему как артефакт (см. Thomas et al. 2022), в отношении этой социотехнической системы могут быть сделаны критериальные, причинные и контекстные заявления.⁷ В то же время, если учёный фокусируется на свойствах и поведении артефакта (например, умного счетчика душа), встроенного в социотехническую систему (например,

умный город), система предоставляет контекст реализации для артефакта, позволяя делать контекстные утверждения об артефакте. Обратите внимание, что артефакт и критерий сравнения существуют в (обычно разных) социотехнических системах, что создает проблемы контекстной валидности.

Чтобы вывести категории критериев, мы рассмотрели основные типы артефактов, созданных наукой дизайна — конструкции, модели, методы, экземпляры и теории дизайна, а также существующие естественные объекты, которые можно использовать в качестве критериев (vom Brocke, Winter и др. 2020; Gregor и Hevner 2013; Hevner и др. 2004; March и Smith 1995). В таблице 3 представлен список критериев, с которыми можно сравнить создаваемый артефакт. Наличие критериев позволяет нам объяснить типы валидности и формализовать конкретную процедуру валидации, которую исследователи могут использовать для поддержки своих заявлений о знаниях.

⁷ После того, как система управления взаимоотношениями с клиентами (CRM) (центральный артефакт) помещена в более широкую социальную и физическую систему организации, мы можем оценить атрибуты более крупной системы, включая организацию (например, удовлетворенность клиентов, справедливость, продажи, прибыль). Кроме того, рассматривая организации с CRM по сравнению с теми, у кого нет CRM, или ту же организацию до и после внедрения CRM, мы можем подтвердить утверждения о влиянии CRM или ее компонентов на интересующие результаты.

Таблица 3. Категории критериев

Категория	Подкатегория	Подробности
Инстанцированные сущности Сущности, генерирующие выход. При использовании они создают изменения в мире, которые можно измерить и записать. Выходы эталонной сущности могут быть в прошлом, настоящем или будущем. Другие аспекты, помимо выходов, также могут сравниваться.	Материальный критерий Реальная сущность, которая становится стандартом, с которым можно сравнивать фокусный артефакт.	Категория включает артефакты критериев (например, поисковая система Google Scholar), объекты реального мира (например, человек-мастер шахмат), процессы реального мира (например, найм сотрудников) или датчики, сообщающие о состояниях реального мира (например, температура или количество осадков). Обязанность исследователя — обосновать, что экземпляр критерия служит разумным стандартом для сравнения. Реализации могут операционализировать конструкции, модели, методы и теории проектирования в материальные решения (например, приложения, платформы, корпоративные системы). Экземпляры критериев используются для демонстрации превосходной производительности артефакта или способности артефакта аппроксимировать выходные данные, структуру или характеристики сущности критерия.
	Манипулируемый артефакт Артефакт, полученный из фокусного артефакта; обычно разрабатывается в рамках того же исследования и обрабатывается для того, чтобы сделать выводы из сравнения двух артефактов.	Создается путем удаления или замены части фокального артефакта.
Неинстанцированные сущности Объекты сравнения, являющиеся чертежами конкретных материальных артефактов . Если эти сущности не инстанцированы, они не производят материальных выходов. Обычно они требуют интерпретации агентом, таким как эксперт-человек.	Теория (и его компоненты: конструкции и принципы проектирования). Система понятий, предназначенная для объяснения, прогнозирования или руководства действиями.	<i>Теории дизайна и их компоненты</i> (например, конструкции, принципы дизайна) и <i>теории, не связанные с дизайном</i> (например, теории объяснения и предсказания), которые могут предоставить знания, имеющие отношение к дизайну для центральных артефактов.
	Модель Концептуальное представление некоторого аспекта реальности, созданное для улучшения понимания или облегчения действия.	Модель представляет собой некоторый аспект реальности; обычно игнорирует аспекты, не имеющие отношения к разработчику модели; может вносить преднамеренные предубеждения. Модель представляет собой область, в которой происходят вмешательства в дизайн или существуют другие артефакты дизайна, и оценивается с точки зрения полезности или соответствия истине.
	Метод Замкнутая, логическая последовательность шагов, используемых для выполнения задачи.	Методы можно найти в самых разных областях. Например, анализ требований, контроль процессов и контрольные списки.

Типы валидности

Типы критериальной валидности: Как указано на рисунке 1, все работы по дизайну должны делать критериальное утверждение, чтобы показать, что спроектированный артефакт обеспечивает некоторую выгоду. Для этого требуется привлечь тип критериальной валидности. Критериальные утверждения (и причинные утверждения, как мы покажем позже) можно сравнивать двумя способами: через их эффективность или через их характеристики. Сравнения эффективности рассматривают сходство выходов артефакта с другими генерирующими выход сущностями, тогда как сравнения характеристик оценивают сходство между характеристиками артефакта и его референтными критериями (таблица 3). Таким образом, критериальная валидность имеет два подтипа: критериально-эффективная валидность и критериально-характеристическая валидность.

Критериально-эффективная валидность поддерживает утверждения о инстанцированных артефактах (таблица 3), когда артефакт и критерий имеют сопоставимые выходы. Критериально-эффективная валидность, таким образом, имеет дело с критериальными утверждениями, которые поддерживаются путем сравнения эффективности артефакта с эффективностью критерия, которая, как утверждается, представляет стандарт. Эти типы валидности поддерживают утверждения о том, что выходы артефакта (или эффекты таких выходов на социотехническую систему) имеют полезность относительно выходов или эффектов критерия.

Можно выделить два подтипа критериально-эффективной валидности в зависимости от того, важно ли время. Если это так, то **прогностическая валидность** с надлежащей регистрацией прогнозов до того, как результаты будут сгенерированы критерием, дает самую сильную валидность для утверждений. Однако в науке дизайна прогностическая валидность чаще

оценивается по будущим данным, которые существуют на момент создания артефакта, но не предоставляются артефакту до окончания валидации. Если время не является фактором в прогнозах, будущие данные недоступны или достаточно более слабой валидации, может быть оценена **текущая валидность**. Когда прогностическое утверждение не сделано, текущая валидность может использоваться для проверки выходных данных артефакта относительно контрольных выходных данных. Часто доступные случаи разделяются с помощью перекрестной проверки на наборы «обучение», «проверка» и «тест», где производительность артефакта относительно истинных значений в наборах «проверка» и «тест» являются типами текущей валидности. В неформальном контексте термин «критериально-эффективная валидность» можно рассматривать как относящийся к текущей валидности.

Критериально – характеристическая валидность имеет дело с критериальными утверждениями, которые сравнивают характеристики артефакта с характеристиками критерия, который, как утверждается, представляет стандарт. Для критериально-характеристической валидности агент (чаще всего человек) обычно участвует в оценке, потому что: эталонный артефакт не производит выходы (как в случае теорий, моделей и методов); выход напрямую не сопоставим с критерием (как это может быть в случае сравнения выходов двух генеративных ИИ); артефакт инстанцирован, но характеристика артефакта (например, его интерфейс) требует субъективной оценки; или артефакт сравнивается с опытом людей с соответствующими артефактами или процессами (например, с оценкой воспринимаемой полезности). Все такие случаи являются случаями, в которых используется критериально-характеристическая валидность.

Критериально-характеристические валидности позволяют оценивать аспекты

артефакта, которые оценки эффективности не могут оценить. Они более гибкие и позволяют сравнивать неидентичные результаты или результаты, которые не могут быть автоматически сравнены (например, восприятие полезности медицинской онтологии или веб-сайта). Независимо от того, основана ли оценка на прямом сравнении результатов или характеристик артефакта, конкретные типы валидности основаны на сущности критерия (таблица 3). Например, валидность экземпляра критерия (валидность экземпляра для краткости) — это степень, в которой поддерживаются утверждения об артефакте относительно материального критерия. Аббаси и др. (2018) занимались обоснованностью экземпляра критерия для проверки утверждения о полезности. Авторы попросили организацию оценить экономию, полученную за счет внедрения системы по сравнению с продолжением текущего бизнес-процесса.

Аналогичным образом, теоретическая валидность касается степени, в которой поддерживаются утверждения об артефакте относительно теоретического артефакта. Например, Лукьяненко и др. (2019) разработали онлайн платформу с конструктивными особенностями, которые, по их утверждению, соответствуют соответствующим принципам из референтной теории дизайна.

Критериальная валидность модели (валидность модели) — это степень, в которой артефакт согласуется с моделью. Хотя этот тип валидности не встречался в выборке рассмотренных нами научных статей по дизайну, рассмотрение критериальной валидности модели присутствовало в нашем обзоре литературы. Например, Рефсгаард и др. (2006, стр. 1596) предположили, что при разработке кода для имитационной модели (артефакт) важно установить, что «код модели в некотором смысле является истинным представлением концептуальной модели» (например, модели, разработанной экспертами) реальной системы (например, экосистемы).

Наконец, критериальная валидность метода критерия (валидность метода) сравнивает метод с существующим методом как самодостаточную логическую последовательность шагов, используемых для выполнения задачи. Валидность метода позволяет оценивать артефакт или часть фокусного артефакта по сравнению с другими критериями (естественными или артефактами и их частями). Пирамуту и Досс (2017) привели пример, когда они оценивали свой артефакт — протокол для одновременной аутентификации нескольких меток радиочастотной идентификации. Они использовали формальное доказательство для подтверждения того, что протокол соответствует самым строгим требованиям безопасности (Avoine et al. 2009).

Типы причинной валидности: Как и критериальная валидность, причинная валидность имеет два подтипа: причинно-эффективная валидность и причинно-характеристичная валидность. Причинно-эффективная валидность касается причинных утверждений, подкрепленных оценкой относительно эффективности измененной версии артефакта, которая имеет намеренно отличающиеся части (по сравнению с проектируемым артефактом). Манипулируемый артефакт обычно разрабатывается в контексте того же исследования той же исследовательской группой и манипулируется, чтобы сделать выводы, вытекающие из сравнений двух артефактов. Иногда это называют исследованием абляции, как это было введено Ньюэллом (1975), и обычно встречается в машинном обучении (например, Аббаси и др. 2012; Аббаси и Чен 2008; Этудо и др. 2017).

Достоверность причинно-следственных характеристик касается причинных утверждений, которые подтверждаются путем сравнения с характеристиками обработанного артефакта, который имеет намеренно отличающиеся компоненты. Один из способов достижения этого различия — удаление или изменение компонента (части). Как и в случае с причинно-эффективной

валидностью критерий обычно разрабатывается в контексте одного и того же исследования одной и той же исследовательской группой и намеренно манипулируется, чтобы сделать выводы из сравнения двух артефактов. Например, Умапати и др. (2008) разработали две версии своего артефакта: одну, позволяющую выбирать шаблоны интеграции с поддержкой речевых актов (основной артефакт), и одну без поддержки (референтный артефакт). Исследователи провели эксперимент, в котором участникам было дано задание по проектированию бизнес-процесса (со списком конкретных элементов задачи) для создания модифицированной модели бизнес-процесса и нотации (BPMN) интеграции предприятия. Модели оценивались экспертами, и было сравнено предполагаемое количество ошибок.

Другой подход основан на контрфактуальном рассуждении (Collins et al. 2004). Сравнение артефакта с версией, в которой удален компонент, может подтвердить, что компонент не оказывает причинно-следственного влияния на производительность артефакта и может быть удален для экономии.

Типы контекстной валидности: контекстная валидность поддерживает утверждения об обстановке и условиях (например, лабораторная или реальная социотехническая система, используемая целевыми пользователями или их доверенными лицами), в которых оценивается артефакт. Цель здесь состоит в том, чтобы оценить, в какой степени критериальные и причинные утверждения, связанные с артефактом, поддерживаются в обстановке, которая в некотором роде похожа на целевую социотехническую обстановку или обобщается в различных обстановках. Следовательно, процедуры оценки контекстной валидности имеют дело с различиями в условиях и обстановках. Контекстные утверждения могут быть сделаны либо против контекста оценки артефакта относительно целевой социотехнической системы, либо против другого контекста. Это приводит к двум подтипам контекстной валидности: экологическая (сосредоточенная на атрибутах исходного контекста оценки относительно атрибутов целевой социотехнической системы) и внешняя (сосредоточенная на оценке в дополнительных социотехнических системах).



Рисунок 2. Структура валидности научного дизайна

Экологическая валидность увеличивается с ростом сходства между социотехнической системой оценки и социотехнической системой, для которой предназначен артефакт. Некоторые подходы к науке дизайна, такие как исследования дизайна действия, как правило, имеют высокую экологическую валидность, поскольку оценки проводятся в натуралистических организационных условиях, в которых используется артефакт. Например, Зайцев и Манкинен (2022), совместно с участниками в Камбодже, работали над созданием приложения, которое можно было бы использовать в реальных условиях.

Внешняя валидность определяется как степень, в которой критериальные или причинные требования к знаниям поддерживаются посредством оценки в более чем одной социотехнической системе. Когда социотехнические системы, в которых оцениваются артефакты, различаются, а артефакт сохраняет свою полезность, реализуется более высокая внешняя валидность. Например, Зайцев и Манкинен (2022) использовали подход исследования дизайна действий для разработки

приложения для повышения финансовой грамотности. Первоначальная разработка произошла в Камбодже, но артефакт был позже адаптирован и использован в Непале, обеспечивая внешнюю валидность, при этом авторы пришли к выводу, что «первоначальный дизайн, уже гибкий и минималистичный, созданный в соответствии с принципами дизайна, обеспечил хорошую основу для локализации» (стр. 106).

Структура валидности науки дизайна и ее применение

Рисунок 2 показывает структуру валидности науки дизайна. Структура устанавливает связь между тремя типами утверждений об создаваемом артефакте и конкретными типами валидности, используемыми для предоставления доказательств в поддержку этих типов утверждений, посредством конкретных типов сравнения.

Таблица 4 формально определяет каждый из типов валидности.

Таблица 4. Определения валидности в науке дизайна	
Валидность	Степень, в которой утверждения о знаниях относительно фокального артефакта подтверждаются доказательствами.
Критериальная валидность	Степень, в которой утверждения о знаниях относительно артефакта подтверждаются путем оценки по сравнению с эталонной сущностью, утверждающей, что она представляет собой стандарт.
Критериально-эффективная валидность	Степень, в которой утверждения о знаниях об эффективности артефакта подтверждаются путем оценки по сравнению с эффективностью материальной эталонной сущности, утверждающей, что она представляет собой стандарт.
Прогностическая валидность	Степень, в которой утверждения об эффективности артефакта подтверждаются путем оценки его точности по сравнению с эталонной эффективностью, которая появилась после того, как данные были использованы для создания артефакта.
Текущая действительность	Степень, в которой утверждения о знаниях об эффективности артефакта подтверждаются путем оценки его эффективности по сравнению с эталонным результатом, появившимся в тот же период, что и данные, использованные для создания артефакта.
Критериально-характеристическая валидность	Степень, в которой утверждения о знаниях о характеристиках артефакта подтверждаются путем оценки его характеристик по сравнению с абстрактными характеристиками эталонной сущности, предположительно представляющей стандарт.
Валидность экземпляра	Степень, в которой утверждения о знаниях о характеристиках артефакта подтверждаются посредством оценки его характеристик по сравнению с

	абстрактными характеристиками инстанцированной эталонной сущности, которая, как утверждается, представляет собой стандарт.
Теоретическая валидность	Степень, в которой утверждения о знаниях о характеристиках артефакта подтверждаются путем оценки его характеристик по сравнению с характеристиками эталонной теоретической сущности, которая, как утверждается, представляет собой стандарт.
Валидность модели	Степень, в которой утверждения о знаниях о характеристиках артефакта подтверждаются путем оценки его характеристик по сравнению с характеристиками эталонной модели, которая, как утверждается, представляет собой стандарт.
Валидность метода	Степень, в которой утверждения о знаниях о характеристиках артефакта подтверждаются путем оценки его характеристик по сравнению с характеристиками эталонной сущности метода, которая, как утверждается, представляет собой стандарт.
Причинная валидность	Степень, в которой утверждения о влиянии части артефакта подтверждаются оценкой по сравнению с обработанным артефактом, в котором эта часть отсутствует.
Причинно-эффективная валидность	Степень, в которой утверждения о влиянии части артефакта на эффективность этого артефакта подтверждаются посредством оценки его эффективности по сравнению с эффективностью измененного артефакта, у которого эта часть отсутствует.
Причинно-характеристичная валидность	Степень, в которой утверждения о знаниях о характеристиках части артефакта подтверждаются посредством оценки его характеристик в сравнении с характеристиками измененного артефакта, у которого эта часть отсутствует.
Контекстная валидность	Степень, в которой утверждения о критериях или причинно-следственных знаниях подтверждаются оценкой в определенном контексте.
Внешняя валидность	Степень, в которой утверждения о критериях или причинных знаниях подтверждаются оценкой в более чем одной социотехнической системе.
Экологическая валидность	Степень, в которой утверждения о критериях или причинных знаниях подтверждаются посредством оценки в социотехнической системе, соответствующей реальным социотехническим системам, для которых предназначен артефакт.

В целом, подходят разные типы валидности в зависимости от видов утверждений, которые хотят сделать исследователи. Как показывает наша структура, существует три основных типа утверждений, соответствующих трем основным типам валидности. Следовательно, наша структура позволяет исследователям определять соответствующие типы валидности на основе характера утверждений о знаниях об артефакте. После того, как исследователи определились с типом утверждения, они должны следовать этому типу утверждения сверху, пока не найдут соответствующий тип валидности нижнего уровня для поддержки этого утверждения.

Например, Koornneef et al. (2020) предложили метод улучшения процесса определения релевантной информации и вариантов решения проблем устранения неполадок/обслуживания воздушных судов между рейсами. Авторы утверждали, что их прототип поддерживал более быстрый поиск

информации, релевантной решению проблем (критериальное утверждение, см. Рисунок 2). Они сравнили свой прототип с текущей практикой в отрасли поиска релевантной информации в руководствах по техническому обслуживанию. Поскольку методы прототипа и ручного поиска давали сопоставимые результаты (т. е. релевантную информацию о проблемах), это пример критериально-эффективной валидности с определенным типом текущей валидности (поскольку предполагается, что время сравнения не играет никакой роли). Чтобы еще больше подкрепить свое утверждение, авторы провели эксперимент, в котором стажеров по техническому обслуживанию попросили найти релевантную информацию, используя либо прототип, либо ручной метод. Авторы измерили время поиска в обоих случаях и обнаружили, что прототип обеспечивал более быстрое время поиска. Если оценка подтверждает утверждение, может потребоваться только один тип валидности.

Если авторы определяют, что их артефакт недостаточно нов, чтобы оправдать публикацию в целевом издании, они могут углубить вклад, сделав причинно-следственное утверждение. Хотя у Koornneef et al. (2020) уже была оценка эффективности,

в гипотетическом сценарии они могли бы затем добавить причинно-эффективную валидность, оценив, какие части их артефакта внесли наибольший вклад в скорость получения релевантной информации.

Таблица 5. Примеры валидации			
Артефакт и тип артефакта	Знание претензии	Тип действия	Оценка
Мет и др. (2015) предлагают инструмент анализа требований — <i>инстанциацию</i> .	Использование инструмента позволяет пользователям определять требования с большей точностью, чем без него.	Текущая валидность	Поэкспериментируйте, чтобы оценить, правильно ли пользователи определяют все требования с помощью инструмента.
Рамакришнан и др. (2023) предлагают теорию <i>принципов проектирования</i> общих знаний на основе платформ.	Принципы обладают доступностью, важностью, новизной и познавательностью, применимостью и направляемостью, а также эффективностью.	Теоретическая валидность	Оценка принципов дизайна в фокус-группе.
Седракан и др. (2017) предлагают метод быстрого прототипирования с обратной связью (FIRP).	Моделирование FIRP улучшает понимание поведенческих аспектов модели.	Валидность модели*	Факторный экспериментальный план, оценивающий понятность метода.
Валеча и др. (2013) предлагают модель общего словаря для стандартизации сообщений в сфере реагирования на чрезвычайные ситуации.	Модель является полной и корректной для удовлетворения потребностей в коммуникации служб экстренного реагирования.	Валидность модели	Фокус-группы с экспертами по чрезвычайным ситуациям.
* Примечание: Критерий может отличаться от типа артефакта, поскольку она занимается установлением базовой линии. Например, участники могут опираться на свою модель мира для оценки некоторых аспектов предлагаемого метода.			

Этот пример иллюстрирует, как наша структура представляет существующие практики науки дизайна, а также как это можно использовать ретроспективно для анализа существующей валидации и лучшего понимания того, почему были сделаны определенные оценки. Она также может предлагать альтернативы (например, оценку прототипов путем запроса человеческого суждения, что приводит к валидностям характеристик), тем самым показывая возможность использования фреймворка для руководства процессом валидации науки дизайна.

Мы далее демонстрируем область применения фреймворка, рассматривая образцы работ из литературы по дизайну. Мы описываем эту демонстрацию, рассматривая типы артефактов, предложенные Марчем и Смитом (1995) и Грегором и Хевнером (2013): инстанции, теория дизайна, методы и модели. Для каждого из этих типов артефактов в Таблице 5 представлен образец статьи и идентифицированы артефакт, утверждения и обоснованность, представленные авторами.

ОЦЕНКА И ПРИМЕНЕНИЕ СТРУКТУРЫ

Чтобы оценить нашу структуру и одновременно продемонстрировать, как её использовать, мы применили фреймворк рекурсивно к самому себе. Фреймворк (тип модели) должен иметь полезность, что означает, например, что он должен рассматриваться как полезный его потенциальными пользователями (Hevner et al. 2004; March and Smith 1995). Фреймворк также должен быть экономичным, при этом точно отражая ключевые концепции домена (National Academies of Sciences 2022).

Мы формулируем несколько утверждений о знаниях о структуре валидности, которые затем проверяем. Мы выдвигаем несколько утверждений о критериях, за которыми следуют утверждения о контексте и причинах, чтобы проиллюстрировать использование всех типов утверждений и различных типов валидности. Во-первых, мы заявляем о полезности структуры, как ее воспринимают потенциальные пользователи — ученые-дизайнеры. Это утверждения о критериях и контекстах, подкрепленные доказательствами *валидности модели* и *экологической валидности*. Во-вторых, мы заявляем о полезности структуры в систематизации существующих *процедур валидации* (названных и неназванных), что подтверждается доказательствами *валидности модели*. В-третьих, мы заявляем о полезности структуры в систематическом сборе и организации существующих типов валидности науки дизайна, как предложено и определено в разных дисциплинах. Это утверждения о критериях и контекстах, подкрепленные доказательствами *причинно-характеристичной валидности*. Наконец, мы утверждаем, что компоненты структуры достаточны для достижения полезности, что подтверждается доказательствами *валидности причинно-следственных характеристик*.

Утверждение 1: Критериальное утверждение и контекстные утверждения с целевыми пользователями

Для обоснования утверждения 1, мы оценили критериальные и контекстные утверждения, проведя две проверки применимости (Rosemann и Vessey 2008), распространенный тип оценки науки дизайна (например, Li et al. 2020; Lukyanenko et al. 2019). Проверка применимости представляет собой оценку, затрагивающую несколько типов валидности. Однако для оценки структуры валидности проверка применимости эффективна для решения вопроса *валидности модели*, когда участники сравнивают нашу структуру со своей моделью идеальной структуры и другими соответствующими структурами. Поскольку она вовлекает вероятных пользователей структуры и просит их использовать ее в контексте своей работы, проверка применимости также предоставляет доказательства *экологической валидности*. Наша оценка проводилась в контексте ученых-дизайнеров, которые имели большой опыт оценки и знакомства с существующими структурами оценки.

Предшествующие оценочные рамки, включая Прата и др. (2015) и Венейбла и др. (2016), фокусируются на моделях оценки и на современных и постфактум стратегиях оценки соответственно. Кодирова шесть измерений методов и критериев оценки, Прат и др. предложили конкретные методы (например, наблюдение, описание, эксперимент, динамический анализ) и конкретные исследовательские подходы (например, качественный, эксперимент, моделирование), которые используются при оценке. Структура Венейбла и др. (2016) фокусируется на искусственном/натуралистическом контексте, в котором происходит оценка, и на том, как эти оценки могут быть формативными (происходящими во время проекта) или суммирующими (происходящими по завершении проекта).

Напротив, наша структура валидности обеспечивает базовое научное обоснование для проверки утверждений о знаниях через процесс оценки. Под этим мы подразумеваем степень, в которой доказательства, полученные в процессе оценки, поддерживают утверждения о знаниях, сделанные исследователями об их артефакте. Более того, наш анализ показывает, что оценки в науке дизайна всегда являются сравнениями: критерии производительности по сравнению с существующими артефактами (или существующими способами достижения того же результата), желаемые характеристики артефакта по сравнению с существующими артефактами или сравнения в пределах и/или между контекстами применения артефакта. При формулировании отношений между критериальной, причинной и контекстной валидностью наша структура валидности выходит за рамки предыдущих структур оценки. Мы показываем процессы, с помощью которых выполняются оценки, и демонстрируем, как устанавливается валидность типов утверждений о знаниях. Делая эти сравнения явными, наша структура предоставляет метакатегории для оценок.

В первой проверке применимости структура валидности была представлена на региональном семинаре по информационным системам, в котором приняли участие как исследователи поведения, так и ученые по дизайну; мы сосредоточились на первоначальной оценке полезности и внесении любых корректировок, прежде чем мы привлечем окончательную когорту экспертов по дизайну. Пятнадцать участников были обучены использованию структуры валидности и получили задание организовать четыре определения валидности и четыре оценки по дизайну в структуру и ответить на открытый опрос. Результаты показали, что структура будет полезна для ученых по дизайну и других исследователей. Например, структура была описана как «вызывающая размышления и освежающая перспектива». В других комментариях отмечалось, что структура была довольно сложной, особенно для

исследователей с ограниченным опытом в области дизайна. В ответ мы добавили процесс выбора типов валидности для использования в исследовательском проекте и оптимизировали структуру.

После включения отзывов в структуру (окончательная версия которой показана на рисунке 2), мы провели вторую проверку применимости среди ведущих в мире ученых-дизайнеров, чтобы оценить важность структуры (критериальное утверждение), доступность (контекстное утверждение) и пригодность (контекстное утверждение).

Одиннадцать участников были учеными-дизайнерами с различным уровнем академического опыта (в среднем 13 лет). Для тех, у кого были учетные записи Google Scholar (девять участников), их среднее количество цитирований составило 4981. Изучение их статей показало, что 36% цитировали Прата и др. (2015) или Венейбла и др. (2016). Проверка применимости проводилась онлайн, чтобы включить ученых из разных областей исследований и подсообществ науки дизайна. Эти участники использовали различные методы, рассматривая различные темы исследований и типы анализа. Их исследования охватывали широкий набор тем, включая моделирование данных и знаний, аналитическое и машинное обучение, науку о данных, изменение бизнес-процессов, моделирование и теорию дизайна. Мы специально пригласили участников из широкого спектра, и чтобы включить участников из науки дизайна в целом. Участники были постоянными участниками WITS, DESRIST и других конференций по науке дизайна. Этапы проверки применимости описаны ниже.

Подготовка: Участникам было предложено описать примеры процессов валидации, которые они использовали в текущем исследовании или недавно завершенном проекте. Эти существующие процессы, часть ментальных моделей экспертов, были эталонами, по которым участники оценивали нашу структуру.

В целом участники продемонстрировали последовательное понимание необходимости оценки и валидации артефактов, но продемонстрировали ограниченное понимание типов валидности, которые соответствовали их оценкам, и не имели словарного запаса для описания своих валидаций. Во многих случаях типы валидности обсуждались в психометрических терминах (например, конструктивная валидность) или в терминах, связанных с полезностью и эффективностью (например, доказательство концепции, доказательство ценности). Большинство участников сосредоточились на процессах, с помощью которых проводились оценки (например, полевые эксперименты, лабораторные эксперименты), а не на роли валидаций в подкреплении заявлений об их артефакте.

Введение в структуру валидности научного дизайна. Затем участники посетили часовую онлайн-сессию, в ходе которой группа авторов представила нашу структуру валидности.⁸ После объяснения структуры и расширенного примера участникам было предложено использовать наш процесс для определения типов валидности в своих проектах.

Результаты. После завершения задания участники заполнили опросник, содержащий шесть открытых вопросов и набор вопросов о полезности артефакта и их намерении использовать, применяя шкалы Лайкерта из Венкатеша и др. (2003). Ответы, как качественные, так и количественные, показали, что участники считали структуру валидности важной, доступной и подходящей (Rosemann and Vessey 2008) для удовлетворения выявленных потребностей сообщества и полагали, что она будет полезна для их собственных исследований.

Роземанн и Весси (2008) определяют *проверку применимости* как исследование,

«которое отвечает потребностям практики, решая реальную проблему» (стр. 3). Процесс оценки этого аспекта является примером критериальной валидности. В нашем контексте ученые-дизайнеры являются практиками в реальном мире, которым самим необходимо подтвердить утверждения об артефактах. Доступность — это критериальная валидность, которая «охватывает то, понятно ли исследование, читаемо ли оно и фокусируется ли оно на результатах, а не на исследовательском процессе» (стр. 3). Наконец, пригодность определяется как степень, в которой исследование может «[удовлетворять] потребностям практики» (стр. 3), что мы понимаем как степень, в которой исследователи рассматривают структуру как подходящую для целевого контекста. Все эти три оценки касаются как модельной валидности, так и контекстной валидности, учитывая их оценку в обстановке, схожей с реальной.

Участники согласились, что структура валидности важна для прояснения все более сложных требований к методам, критериям и стратегиям валидации. Эта важность была подтверждена такими комментариями, как «она поддерживает систематический подход к валидации» и «такой фреймворк может дать исследователям принятый стандарт, на который можно ссылаться, когда они пытаются оценивать свои собственные артефакты науки дизайна». Другие комментарии признали, что фреймворк снижает сложность, связывает утверждения о знаниях со стратегиями валидации и предоставляет детали, полезные для разработки утверждений о валидности. Все эти комментарии поддерживают наше критериальное утверждение.

Несмотря на то, что было только краткое введение в структуру, доступность была подтверждена такими комментариями как: «Хорошо, что я узнал о вашей структуре валидности на раннем этапе моего проекта». Один из участников нашел структуру доступной сразу, заявив: «Я могу использовать структуру структура

⁸ Презентация доступна в прозрачных материалах по адресу <https://osf.io/ca6vg/>.

предоставила подтверждение того, что моя оценка артефакта достаточно хороша, но мне следовало бы разработать несколько дополнительных деталей в своей оценке».

Окончательный предшественник применимости, пригодность, также была отражена в комментариях: «Как ученый и автор DSR, структура помогает мне планировать проверку моей работы на ранней стадии. ... Как рецензент и редактор, я смогу указать на общее понимание валидности в DSR. Оценивая ценность статьи DSR, я могу использовать структуру для выявления сильных сторон и потенциала для улучшения. Или побудить авторов поделиться аспектами своей работы DSR, которые они исключили из статьи, но которые имеют отношение к установлению ее валидности».

Были высказаны два опасения относительно будущих последствий структуры валидности. Во-первых, структура валидности может быть использована для сдерживания публикаций, поскольку «рецензенты могут не согласиться с тем, какие типы валидности требуются». Участники согласились, что это не недостаток самой структуры валидности. В ответ, мы уточняем, что нецелесообразно использовать фреймворк в качестве аргумента для дополнительных оценок. Как отмечалось выше, дополнительные оценки оправданы только в том случае, если либо выполненные оценки не соответствуют заявленным знаниям, либо вклад статьи в существующие заявления о знаниях считается группой рецензентов недостаточным. В последнем случае любые дополнительные оценки, предлагаемые рецензентами, должны соответствовать дополнительным заявлениям о знаниях.

Вторая проблема заключалась в потенциальном увеличении затрат и времени на выполнение и проверку проектно-конструкторских работ из-за растущих требований к процедурам проверки валидности. Мы обновили нашу первоначальную структуру, чтобы лучше объяснить утверждения о знаниях и процесс

применения структуры для защиты от такого использования структуры.

Участники ответили по 7-балльной шкале Лайкерта о полезности структуры валидности ($\mu = 5,64$; $\sigma = 1,04$), распространенный вопрос проверки применимости в науке дизайна (например, Ли и др. 2020; Лукьяненко и др. 2019). Их также опрашивали об их намерении использовать структуру ($\mu = 6,15$; $\sigma = .92$) относительно их существующего процесса, еще одно критериальное утверждение, которое при оценке в целевой установке удваивается как контекстное утверждение. Что касается полезности, один участник отражал вторую обеспокоенность по поводу того, замедлит ли структура (по крайней мере, изначально) выполнение исследовательских задач. Без этого вопроса участника полезность заметно возросла (с $\mu = 5,64$ до $\mu = 6,14$; $\sigma = 1,03$).

Один из участников, который беспокоился, что структура валидности может быть сложной в использовании, в то же самое время намеревался её использовать, заявив, что «существование такой структуры может дать исследователям общепринятый стандарт, на который можно сослаться, когда они пытаются подтвердить свои собственные артефакты [заявления] в области науки дизайна». Все участники указали, что они намерены использовать структуру, как только она станет доступной. Поэтому мы пришли к выводу, что эти ответы показывают *валидность модели* и добавляют доказательства *экологической валидности*.

Утверждение 2: Критериальное утверждение полноты структуры

Мы теперь оцениваем утверждение о том, что структура валидности является полной в представлении существующих оценок в опубликованных статьях по дизайну. Поскольку структура является моделью, она должна быть в состоянии представлять эти оценки, обеспечивая *валидность модели*. Мы проконсультировались с двумя источниками, чтобы определить совокупность

соответствующих статей, по которым можно было бы сделать выборку. Во-первых, мы использовали 121 статью по дизайну, идентифицированную Пратом и др. (2015) из Ассоциации по информационным системам (AIS) Senior Scholars' Basket of Eight journals (с апреля 2004 г. по декабрь 2013 г.). Во-вторых, мы проанализировали 1233 дополнительных статьи-кандидата из AIS Basket of Eight с 2014 по 2017 год. Затем мы вручную закодировали эти 1233 статьи на основе схемы включения Прата и др. (2015). Восемьдесят шесть статей были закодированы третьим автором как наука дизайна и перепроверены вторым автором. Было достигнуто 100% согласие, что все были статьями по науке дизайна. Исходные 121 статья от Прата и др. (2015) плюс наши 86 статей дали в общей сложности 207 статей, опубликованных в период с апреля 2004 года по декабрь 2017 года.

Чтобы охватить статьи, которые следовали подходу науки дизайна, но явно не использовали фразу «наука дизайна», мы создали наборы признаков на основе списка ключевых слов науки дизайна, а также цитат из лучших научных статей по дизайну. Затем мы обучили модель машинного обучения для различения соответствующих рукописей в полном наборе рукописей на основе процесса, описанного Ларсеном и др. (2019). Мы применили комбинацию машинного обучения и ручной оценки, чтобы расширить годы охвата с 1994 по 2019 год и источники, включив AIS Basket of Eight, а также *Decision Support Systems*, *ACM Transactions on Information Systems* и труды Международной конференции по научным исследованиям в области дизайна в области информационных систем и технологий (DESRIST) и Международной конференции по информационным системам (ICIS). После случайной выборки из последних четырех источников мы получили 527 научных статей по дизайну.

Из нашего корпуса мы случайным образом отобрали 32 эмпирические статьи, которые четвертый и шестой авторы независимо закодировали, выделив 79 описаний оценки.

Затем первый и третий авторы независимо закодировали каждое описание оценки, используя структуру. Кодировщики согласились в 76,3% случаев, что привело к капле Коэна 0,703. Разногласия в основном касались того, что намеревались утверждать первоначальные авторы, а не того, как интерпретировать структуру валидности. Из этого упражнения мы пришли к выводу, что оценки, найденные в рассмотренных работах, можно классифицировать с использованием типов валидности в нашей структуре, демонстрируя *валидность модели*. Однако два типа валидности, *теоретическая валидность* и *валидность модели*, не использовались в отобранных нами работах (таблица 6).

Таблица 6. Определения достоверности (затененные ячейки означают «неприменимо»)

Имя	Часть	Выход	Артефакт	СТС
Валидность				
Критериальная валидность				
Критериально-эффективная валидность			19	
Прогностическая валидность			3	
Текущая валидность			4	
Критериально-характеристичная валидность				
Валидность экземпляра	0	0	5	
Теоретическая валидность	0	0	0	
Валидность модели	0	0	0	
Валидность метода	0	0	10	
Причинная валидность				
Причинно-эффективная валидность	24			
Причинно-характеристичная валидность	7	0	1	
Контекстная валидность				
Внешняя валидность				7
Экологическая валидность				1

Утверждение 3: Критериальные и контекстные утверждения

Формализация общих дисциплинарных норм ускоряет научный прогресс (Национальные академии наук 2022). Многие дисциплины разрабатывают и проверяют артефакты, и структура валидности должна быть способна представлять проверки, используемые для оценки этих артефактов.

Утверждение 3 заключается в том, что структура валидности является полной в представлении типов валидности науки дизайна не только для информационных систем, но и для поведенческой науки, инженерии и медицины — когда эти дисциплины подтверждают утверждения о своих артефактах. Оценивая утверждения о том, что структура является полной в своей способности представлять определения валидности науки дизайна, мы устанавливаем *валидность модели*. Показывая, что наша структура применима к другим дисциплинам, мы устанавливаем *внешнюю валидность*.

Первой задачей было определить конкретные типы валидности в области науки дизайна, уже предложенные в литературе. До нас не было существующего набора общих типов валидности. Мы создали такой набор для дальнейшего уточнения подмножества типов валидности. Мы начали с определения и сбора типов валидности и определений из различных областей, включая социальные науки, инженерию и информатику, а также медицину. Это было составлено в течение трехлетнего периода первым автором и группой научных сотрудников. *Первоначальные источники* были документы, содержащие наборы типов валидности, такие как стандарты для образовательного и психологического тестирования (например, Американская ассоциация образовательных исследований и др. 2014). Самая большая коллекция названий валидности, 168, была предоставлена Ньютоном и Шоу (2014). Подавляющее большинство из них было в большой таблице как демонстрация неразрешимой природы типов валидности. Тем не менее, многие из них не появились ни в одной доступной базе данных и не могли быть найдены с помощью поисковой системы.

Вторая задача заключалась в получении определений для выявленных типов валидности. Мы в основном полагались на научные книги и статьи. Для каждого

источника раздел, содержащий валидность, был дополнительно изучен для выявления *валидности-кандидатов*. Валидность-кандидат представляла собой концепцию, которую автор либо называл валидностью, либо концепцию, не названную таковой, но указанную с типами валидности (например, *обыденный реализм* как тесно связанный с *экологической валидностью*). Мы не подвергали сомнению утверждения авторов. В каждом случае мы записывали новые типы валидности-кандидатов. Всего из приблизительно 7500 вручную проверенных источников было получено 2418 типов валидности-кандидатов. Они были категоризированы по названию, прежде чем каждый набор категорий был изучен для поиска омонимов (одинаковое название, но разное значение), что дало 418 различных типов валидности. Этот обзор литературы был использован для создания первой версии нашей структуры.

В третьей задаче мы изучили все статьи, опубликованные в журнале *AIS Basket of Eight* (Lowry et al. 2013) с 1994 по 2017 год. Всего было проанализировано 6083 статьи с применением 216 запросов регулярных выражений, представляющих валидность и связанные с ними концепции, такие как обобщаемость, что дало 73 365 предложений. Предложения были упорядочены по количеству совпадений, при этом 9707 предложений содержали более двух совпадений по регулярным выражениям, вручную проверенным первым автором. Этот анализ дал только 23 дополнительных типа валидности, что в общей сложности составило 441 типов валидности, что говорит о том, что изначальный процесс был тщательным. Мы исключили из дальнейшей оценки любой потенциальный тип валидности, не являющийся общеупотребительным, для которого нам не удалось найти пять определений из разных источников. На этом этапе было удалено сто пятьдесят восемь потенциальных типов валидности.

В ходе четвертого задания первый и второй авторы независимо друг от друга прочитали

пять определений для каждого типа валидности и выбрали одно или два определения, которые представляли общие аспекты, выраженные другими определениями. Решения обсуждались (и записывались), и разногласия разрешались для достижения соглашения по одному или двум определениям, представляющим тип валидности. Поскольку было много случаев, когда существовали очень похожие или даже идентичные типы валидности, расчет показателей межэкспертной надежности был нецелесообразным. Одиннадцать потенциальных типов валидности были исключены на этом этапе, поскольку они не соответствовали нашему определению исследовательской валидности (например, валидности, основанные на законах) или поскольку четкого определения не появилось.⁹

Для пятой задачи, с окончательным набором указанных типов валидности, первый и второй авторы следовали нашему установленному определению валидности науки дизайна и независимо кодировали все типы валидности как «тип валидности науки дизайна» или «другой тип валидности». Кодировщики достигли 90,4% согласия с каппой Коэна 0,79. Кодировщики обсудили и разрешили все разногласия. Почти все разногласия возникли из-за типов валидности эффективности, где один кодировщик использовал более инклюзивную интерпретацию. Всего было обнаружено 79 типов валидности науки дизайна.

Из типов валидности науки дизайна мы удалили типы валидности, которые объединяли несколько других типов валидности. После удаления этих типов валидности осталось 70 определений для категоризации в нашей структуре. Из них 23 (33%) поступили из поведенческих научных изданий, 37 (53%) — из инженерии и компьютерных наук и 10 (14%) — из

⁹ Обоснование каждого удаления доступно в прозрачных материалах.

медицины; 22 (31%) были взяты из выборки журналов IS.

Первый и третий авторы независимо друг от друга закодировали все определения валидности в фреймворке и пришли к согласию в 79,2% случаев (каппа Коэна = .731), включая случаи частичного согласия, закодированные как несогласие (например, валидность эффективности критерия против прогностической валидности). Разногласия обсуждались и разрешались. Все определения вписываются в категорию валидности фреймворка (левый столбец Таблицы 7).¹⁰

¹⁰ Примечание: метрики валидности были включены в анализ, но, поскольку они не используют утверждение, они сами по себе не являются валидностями. Поэтому мы исключили из классификации следующие метрики: точность, площадь под кривой, полнота, правильное отклонение, уровень обнаружения, F1-оценка, выпадение, ложная тревога, доля ложного обнаружения, ложный отрицательный результат, уровень ложного упущения, ложный положительный результат, уровень ложного положительного результата, попадание, уровень попадания, коэффициент корреляции Мэтьюза, промах, уровень промахов, отрицательная прогностическая ценность, положительная прогностическая ценность, точность, отзыв, чувствительность, специфичность, истинно отрицательный результат, истинно отрицательный результат, истинно положительный результат, истинно положительный результат.

Таблица 7. Кодирование определения валидности	
Научная обоснованность дизайна	Валидности в литературе
Критериальная валидность	
Критериально- эффективная валидность	Абсолютная валидность, валидность группы критериев, валидность критерия, валидность, ориентированная на критерии, валидность, связанная с критериями, валидность решения, диагностическая валидность, дискриминативная валидность, эмпирическая валидность, валидность низшего порядка, операциональная валидность, прагматическая валидность, процедурная валидность, постдиктивная валидность, репликативная валидность, ретроспективная валидность, валидность приложения
Прогностическая валидность	Прогностическая валидность критерия, прогностическая валидность (2)*, прогностическая критериальная валидность
Текущая валидность	Совпадающая критериальная валидность, совпадающая валидность, перекрестная валидность, относительная валидность
Критериально-характеристичная валидность	
Валидность экземпляра	Наблюдательная валидность, физическая валидность
Теоретическая валидность	Этиологическая валидность, теоретическая валидность, валидность реализации
Валидность модели	Концептуальная валидность модели, функциональная валидность, структурная валидность, семантическая валидность
Валидность метода	Алгоритмическая валидность, последовательность
Причинная валидность	
Причинно- эффективная валидность	
Причинно- характеристичная валидность	
Контекстная валидность	
Внешняя валидность	Прагматическая обоснованность
Экологическая валидность	Поведенческая валидность, экологическая валидность
Примечание: *две немного разные версии прогностической валидности, ссылающиеся на один и тот же тип валидности.	

Таким образом, наша оценка дала веские доказательства обоснованности утверждения 3 с точки зрения как критериального утверждения, так и контекстного утверждения, усиливая *модельную валидность* и обеспечивая *внешнюю валидность*. Она не подтвердила необходимость двух причинных валидностей, включенных в структуру. Отсутствие внимания к причинной валидности в существующих определениях типов валидности предполагает возможность внести вклад в обоснованность науки дизайна. В этом смысле она подтверждает необходимость шаблонов в науке дизайна, как предложено Пейфферсом и др. (2008).

Утверждение 4: Причинное утверждение о том, что структура является экономной

Структура валидности науки дизайна была разработана поэтапно, как это принято в науке дизайна. Поскольку мы хотели оставаться инклюзивными, первоначальная версия структуры была более обширной, чем та, которая представлена в этой статье. Например, первоначальная структура содержала тип валидности, называемый «валидность требования» — тип критериально-характеристичной валидности, используемый при сравнении артефакта с документом требований или ожиданиями и опытом пользователя. Однако требования не являются артефактами, которые отражают

вклад в науку и общество, и мы не смогли четко классифицировать валидность требований как отдельный тип валидности. Поэтому мы занялись оценкой причинно-характеристичной валидности, чтобы изучить необходимость в этом типе валидности (поддерживая утверждение, что наша структура без этого типа валидности является экономной и не имеет больше компонентов, чем необходимо).

Валидность требований была введена для рассмотрения претензий на знания в отношении явных требований (например, функции, простоты использования, формы) и неявных требований (например, потребностей, целей или опыта с аналогичным классом артефактов). *Требование* является распространенным артефактом, создаваемым при разработке программного обеспечения. Тем не менее, оно не фигурирует в общих списках типов артефактов науки дизайна (например, Gregor и Nevner 2013; March и Smith 1995). Однако изначально мы считали его одним из критериев нашей структуры, склоняясь к большей инклюзивности и консервативности. Поскольку наше понимание области валидности со временем углублялось, мы начали задаваться вопросом, является ли *требование* необходимым. Мы решили решить эту проблему, сделав формальное заявление о том, что *требование* является необходимым типом валидности для полноты фреймворка.

Для проверки этого утверждения мы выбрали подход к оценке, основанный на контрфактуальном рассуждении. Первый и третий авторы совместно закодировали два определения валидности, ранее закодированные как «валидность требования» в исходной структуре. Во-первых, мы изучили «валидность приложения», определяемую как соответствие имитационной модели ее цели и требованиям; т. е. вероятность того, что модель выдает результаты, которые отражают некоторый внешний артефакт или социотехническую систему. На основе нашего улучшенного понимания структуры,

этот экземпляр валидности был перекодирован как критериально-эффективная валидность. Вторым существующим определением была «семантическая валидность», определенная частично на основе уместности определений категорий. Мы поняли, что критерием было понимание семантическими оценщиками эквивалентной модели или категории социотехнической системы, предполагая, что это была *валидность модели*. Та же логика стала яснее при изучении семи оценок науки дизайна, изначально отнесенных к категории «валидности требования». В большинстве случаев первоначальная оценка фокусировалась на таких терминах, как «требование» или «удовлетворение» и «приемлемость». Часто оценки были плохо описаны первоначальными авторами и неясны с точки зрения фактического эталонного артефакта или социотехнической системы. В этой оценке причинной валидности мы сосредоточились на том, что, по нашему мнению, исследователи считали своим эталонным артефактом. Для семи оценок мы обнаружили пять случаев *модельной валидности* и два случая *экземплярной валидности*.

Таким образом, эта оценка показала, что фреймворк валидности можно сделать более экономичным без потери репрезентативной способности. Наша контрфактуальная оценка *причинной валидности* продемонстрировала, что удаленные части фреймворка не были каузально вовлечены в производительность фреймворка. Все другие типы валидности были построены вокруг общепризнанных типов артефактов и были необходимы для классификации доказательств, демонстрируя *причинно-характеристичную валидность* для первоначального утверждения об экономичности для оставшихся частей фреймворка. Поэтому маловероятно, что дополнительные оценки причинной валидности позволят удалить дополнительные части. В результате, дальнейшие итерации не были сочтены необходимыми.

ОБСУЖДЕНИЕ И ПОСЛЕДСТВИЯ

Общий вклад

Критическая часть науки дизайна — это проверка утверждений о знаниях о создаваемых артефактах. Это исследование формулирует, почему необходимы процедуры проверки, и предоставляет процесс, с помощью которого они могут быть идентифицированы и реализованы с помощью различных типов валидности.

Мы создали Design Science Validity Framework который сопоставляет типы валидности с характеристиками утверждений о знаниях, помогая авторам формулировать и сообщать свои утверждения о знаниях и доказательства, их подтверждающие. Основанный на самом большом обзоре литературы о валидности, Design Science Validity Framework представляет собой всеобъемлющую структуру для валидации заявлений о знаниях об артефактах. Структура предоставляет стандартный словарь для отчетности о валидности исследований. Сама структура имеет внешнюю валидность, поскольку успешно представляет все идентифицированные доказательства валидности, связанные с артефактами, из информационных систем, поведенческой науки, инженерии, компьютерных наук и медицины.

Структура направляет исследователей в определении утверждений о знаниях об артефактах, рассматривая ветви иерархии типов валидности, тем самым усиливая строгость и вклад проектов в области науки дизайна. Предоставляя процесс и номенклатуру для проверки утверждений о знаниях, структура может применяться к любому типу создания и оценки артефактов. Явное связывание типов валидности, процессов оценки и подтверждающих доказательств с утверждениями о знаниях должно быть полезным для исследователей и рецензентов в области науки дизайна, а также для тех, кто принимает другие исследовательские подходы. Мы уверены что, наша структура может помочь наладить

связи между наукой дизайна и другими типами исследований. Развернутые утверждения о знаниях также могут использоваться для расширения вкладов в предшествующую работу. Это можно сделать, например, сделав причинные и контекстные утверждения об артефактах или добавив дополнительные критериальные утверждения с другими критериями.

Выводы

Построение и оценка Design Science Validity Framework привели к нескольким примечательным идеям. Во-первых, ученые-дизайнеры исторически использовали разные оценки для поддержки утверждений о валидности самого артефакта. Чтобы определить валидность как аспект качества в науке дизайна, мы переносим эти оценки на установление валидности явных утверждений о знаниях об артефакте. Эти утверждения будут иметь различную степень подтверждающих доказательств, достаточность которых будет установлена сообществом и потенциально оспорена и изменена с течением времени.

Во-вторых имеется, на удивление мало обсуждений валидности науки дизайна, хотя оценки и типы валидности хорошо известны и понятны в других дисциплинах. На данный момент наука дизайна использует узкий диапазон концепций валидности, в основном фокусируясь на мерах эффективности, что говорит о том, что валидность недоиспользуется. В нашем обзоре литературы наиболее часто встречающимися терминами были точность, отзыв, специфичность, истинно положительный и ложноположительный (концепции, связанные с матрицей путаницы). Это, по сути, метрики, а не типы валидности, но они используются при установлении валидности эффективности. В выборке проанализированных публикаций эти метрики валидности чаще всего были связаны с оценкой моделей машинного обучения, что, конечно, не отражает всю сферу науки дизайна.

В-третьих, некоторые типы валидности, используемые в науке дизайна, были заимствованы из других дисциплин. Однако это не способствует целостному оценочному подходу для науки дизайна. Например, валидность в психометрии и эконометрике в значительной степени сосредоточена на измерении, а некоторые типы полезны в науке дизайна (например, внутренняя валидность для поддержки утверждений о причинности в экспериментах). Однако психометрические валидности недостаточны для поддержки диапазона утверждений о знаниях в науке дизайна.

Наконец, как подтвердила наша проверка применимости, основной проблемой для исследователей, создающих артефакты, является неопределенная природа валидации и тенденция выполнять валидационные действия неявно, с небольшим пониманием базовой структуры таких валидаций. Структура валидности предоставляет исследователям структурированный шаблон и набор тщательно объясненных типов валидности с явными стандартизированными определениями. Процесс валидации позволяет исследователям собирать доказательства, которые поддерживают тип валидности, явно связанный с утверждением о знаниях о разработанном артефакте.

Практические предложения

Обобщая аргументы и выводы нашей статьи, мы даем следующие рекомендации исследователям, заинтересованным в разработке артефактов, которые вносят вклад в науку и общество.

Чётко делайте утверждения об артефактах. Чтобы достичь высокого уровня практической полезности и научной воспроизводимости, исследователи должны делать точные и явные утверждения о своих артефактах. Это должно помочь сформулировать исследовательский вклад и направить валидацию, поскольку валидация зависит от сформулированных утверждений.

Учитывая, что знания о дизайне могут развиваться посредством поэтапного улучшения артефакта (Tuunanen et al. 2024), утверждения могут возникать на разных этапах. Например, некоторые утверждения могут быть сделаны до разработки артефакта, в то время как другие утверждения могут быть сделаны после развертывания артефакта в некотором контексте и наблюдения за результатами. Однако последние утверждения не следует считать подтвержденными, пока они не пройдут соответствующую процедуру проверки, возможно, в последующей итерации артефакта. Важно то, что утверждение о знаниях формально оценивается независимо от процесса, который сформировал утверждение, что требует, чтобы утверждения были заявлены явно.

Делайте утверждения соразмерные предполагаемому вкладу. Структура валидности не должна использоваться для оправдания чрезмерных валидаций. Часто невозможно/нужно заявлять каждое утверждение об артефакте, и не все утверждения могут быть подвергнуты валидации (например, из-за сложности проведения сравнений или получения подходящих сравнительных критериев). Вопрос о том, сколько доказательств необходимо, зависит от контекста и конкретных характеристик проблемы.

Исследователи должны сделать по крайней мере одно критериальное утверждение об артефакте, стремясь провести сравнение с современными артефактами или процессами. Этого может быть достаточно, если артефакт особенно новый, так что мало что известно о том, что делает его эффективным, или о дополнительных (помимо исходных) контекстов, где он может быть развернут. Помимо этого, причинные и контекстные утверждения усиливают исследовательский вклад, поскольку они углубляют знания о дизайне и помогают практикам надежно и безопасно применять результаты исследования в различных средах.

Убедитесь, что каждое утверждение о знаниях подтверждено. Если изначальное утверждение сформулировано об артефакте, его следует подтвердить. Например, если утверждается, что какой-то компонент артефакта вызывает определенный результат, то установление причинной валидности является уместным. Хотя единичная проверка не доказывает утверждение о знаниях, подтверждение утверждений увеличивает вероятность получения надежных знаний о дизайне.

Что касается того, какой объем оценки необходим для установления обоснованности, Галисон и Д'Агостино (1987) представили убедительный аргумент о том, что достаточность доказательств является вопросом общественного согласия. Количество требуемых действий по валидации обычно основано не на конкретном правиле, а на консенсусе относительно набора «убедительных аргументов, которые будут «выступать в суде»» (Galison and D'Agostino 1987, стр. 227). Исследователи и группы рецензентов достигают такого консенсуса в процессе рецензирования. Важно то, чтобы применяемые процедуры валидации были подходящими для валидации заявлений о знаниях, сделанных в статье.

Проверяйте утверждения на протяжении всего пути артефакта. Многие артефакты появляются в ходе многоэтапных процессов посредством экспериментов, доработки или постепенного улучшения. Некоторые проекты включают несколько этапов и производят различные артефакты на этих этапах (например, концептуальная модель, прототип системы, бета-версия, окончательная система в производстве) (Tuunanen et al. 2024). Утверждения о знаниях о полученных артефактах могут быть сделаны на протяжении всего этого процесса. Формирующая промежуточная проверка этих утверждений может помочь в дальнейшем уточнении и улучшении артефакта. «Промежуточные» проверки могут быть неформальными, например, с использованием удобной выборки

потенциальных пользователей или легкодоступного критерия. Однако для того, чтобы гарантировать, что полученный артефакт надежно способствует науке и практике, утверждения о *финальной* или *публичной* версии артефакта должны подвергаться строгой суммирующей финальной проверке.

Убедитесь, что используются соответствующие типы валидности. С созданием Design Science Validity Framework исследователь может ссылаться на структуру во время проверки своих заявлений о знаниях. Структура организует различные практики валидности в согласованную систему ссылок. Она показывает, какие типы валидности подходят для каждого типа утверждения, и предлагает сравнительные критерии и процедуры сравнения, разумные для этих типов валидности. Поскольку научное сообщество продолжает применять и совершенствовать эти типы валидности, ожидается, что их надежность будет расти, предоставляя исследователям более прочную основу для построения своих исследований.

Структура валидности в дизайне предоставляет возможности для будущих исследований. Во-первых, структура является расширяемой, что означает, что она может включать дополнительные подтипы валидности, которые могут быть предложены исследовательским сообществом (например, дальнейшее уточнение валидности модели на основе типов моделей). Во-вторых, можно лучше отслеживать закономерности валидации в дизайне и выявлять пробелы и возможности (например, потребность в большем количестве контекстных утверждений). Связанная с этим возможность заключается в улучшении процедур валидации путем обеспечения того, чтобы соответствующие практики валидации систематически фиксировались для соответствующих типов валидности. Наконец, исследователи могут исследовать применение структуры в проектах по дизайну и сообщать о результатах, связанных с полезностью и долгосрочным влиянием использования структуры на зрелость

дизайна и его интеграцию с другими исследовательскими традициями.

ЗАКЛЮЧЕНИЕ

В этом исследовании определяется валидность для науки дизайна и предлагается структура валидности науки дизайна и процесс ее использования. Структура, полученная в результате обширного обзора и анализа литературы по валидности, определяет и организует неявные и явные типы валидности в три основные категории: *критериальная, причинная и контекстная*. Категории валидности предназначены для того, чтобы помочь исследователям в строгом получении и представлении доказательств своих утверждений о знаниях. Отображение утверждений о знаниях в структуру валидности должно поддерживать ученых и помогать улучшать научные знания, связанные с артефактами. Сама структура была валидирована путем оценки ее собственных утверждений о знаниях и предоставления доказательств для соответствующих типов валидности, которые поддерживают эти утверждения.

Благодарности

Авторы хотели бы поблагодарить старшего редактора Баладжи Падманабхана, помощника редактора Стефана Сайделя и анонимных рецензентов за их работу над предыдущими версиями нашей статьи. Мы ценим отзывы, полученные от участников семинаров в HICSS и ECIS, а также от Университета Колорадо, Университета Аризоны, Университета штата Аризона и Сиднейского университета.

Финансирование. Это исследование было частично поддержано финансированием от Национального института здравоохранения США, предоставленным Каю Р. Ларсену [Грант 3U24AG052175-08S1], а также финансированием от Совета по естественным наукам и технике Канады Джеффри Парсонсу [Грант RGPIN-2020-04916].

References

- 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*, Amer 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 *DESIRIST*, 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 Assessment," *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 *DESIRIST 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.
- Ethayarajh, 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, UK.
- 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.
- Peffers, K., Tuunanen, T., and Niehaves, B. 2018. “Design Science Research Genres: Introduction to the Special Issue on Exemplars and Criteria for Applicable Design Science Research,” *European Journal of Information Systems* (27:2), pp. 129–139.
- Peffers, 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.

Ссылки на русском языке

- Аббаси, А., Альбрехт, К., Вэнс, А. и Хансен, Дж. 2012. «Метамошенничество: структура метаобучения для обнаружения финансового мошенничества», *MIS Quarterly*, стр. 1293–1327.
- Аббаси, А. и Чен, Х. 2008. «CyberGate: структура проектирования и система для анализа текста компьютерно-опосредованной коммуникации», *MIS Quarterly* (32:4), стр. 1–30.
- Аббаси, А., Чжоу, И., Дэн, С. и Чжан, П. 2018. «Текстовая аналитика для поддержки осмысления в социальных сетях: перспектива языка и действия», *MIS Quarterly* (42:2), стр. 1–38.
- Американская ассоциация образовательных исследований, Американская психологическая ассоциация, Национальный совет по измерениям в образовании, Объединенный комитет по стандартам образовательного и психологического тестирования (США). 2014. *Стандарты образовательного и психологического тестирования*, Американская ассоциация образовательных исследований.
- АРА. 2020. «Validity», Американская психологическая ассоциация. (<https://dictionary.apa.org/validity>).
- Арази, О., Кумар, Н. и Шапира, Б. 2010. «Основа теории проектирования для систем социальных рекомендаций», *Журнал Ассоциации информационных систем* (11:9), стр. 455–490.
- Авдиджи, Х. и Винтер, Р. 2019. «Пробелы в знаниях в исследованиях в области дизайна», в *DESRIST*, стр. 1–9.
- Авуан, Г., Бингол, МА, Кардас, С., Лораду, К. и Мартин, Б. 2009. «Формальная структура для криптоанализа протоколов ограничения расстояния Rfid». *Эта работа частично финансируется проектом FP7-ICE в рамках соглашения о гранте (206546)*, Citeseer.
- Баскервиль, Р., Кауль, М. и Стори, В. К. 2015. «Жанры исследований в области дизайна и науки: обоснование и оценка производства знаний», *MIS Quarterly* (39:3), стр. 541–564.
- Будро, М.-К., Гефен, Д. и Штрауб, Д.В. 2001. «Валидация в исследованиях информационных систем: современная оценка», *MIS Quarterly* (25:1), стр. 1–16.
- von Brocke, J., Gau, M. и Mädche, A. 2021. «Журналирование процесса исследования науки о дизайне. Прозрачность в создании знаний о дизайне», в *DESRIST 2021*, Springer, стр. 131–136.
- фон Брокке Дж., Хевнер А.Р. и Маедче А. 2020. *Исследования в области науки о дизайне: примеры*, Берлин / Гейдельберг: Springer.
- фон Брок, Дж., Винтер, Р., Хевнер, А. и Мэдче, А. 2020. «Накопление и эволюция знаний о дизайне в исследованиях в области дизайна: путешествие во времени и пространстве», *Журнал Ассоциации информационных систем*.
- Бертон-Джонс, А., Бох, У. Ф., Оборн, Э. и Падманабхан, Б. 2021. «Комментарии редактора: Повышение прозрачности исследований в MIS Quarterly: плюралистический подход», *MIS Quarterly* (45:2), стр. iii–xviii.
- Картер, Дж. В. 2019. *Аристократ о ранней греческой психологии: наука о душе*, Кембридж, Великобритания: Cambridge University Press.
- Чан, Э.К. 2014. «Стандарты и рекомендации по практике валидации: разработка и оценка измерительных приборов», в книге «Валидность и валидация в социальных, поведенческих и медицинских науках», Springer, стр. 9–24.
- Чаттерджи, С., Саркер, С., Ли, М.Дж., Сяо, Х. и Элбанна, А. 2021. «Возможная концептуализация артефакта информационных систем (ИС): перспектива общей теории систем 1», *Журнал информационных систем* (31:4), Онлайн-библиотека Wiley, стр. 550–578.
- Чоудари, Г. Дж. и Канхангад, В. 2022. «Двухветвевая сеть для диагностики заболеваний грудной клетки с помощью рентгенографии грудной клетки», *Журнал IEEE по биомедицинской и медицинской информатике* (26:12), IEEE, стр. 6081–6092.
- Коэн, Л., Манион, Л. и Моррисон, К. 2013. *Методы исследования в образовании*, Лондон, Англия: Routledge.
- Коллиер-Рид, Б. и Ингерман, А. 2013. «Феноменология: от критических аспектов к притязаниям на знания», в «Теории и методе в исследованиях высшего образования» (т. 9), Emerald Group Publishing Limited, стр. 243–260.
- Коллинз, Дж., Холл, Н. и Пол, Л.А. 2004. *Причинно-следственная связь и контрфактуальности*, Кембридж, Массачусетс: MIT Press.
- Кук, Т.Д. и Кэмпбелл, Д.Т. 1979. *Квазиэксперимент: вопросы проектирования и анализа для полевых*

- условий , Чикаго: Rand McNally College Pub. Co.
- Кресвелл, Дж. В. и Миллер, Д. Л. 2000. «Определение валидности в качественном исследовании», *Теория в практику* (39:3), Тейлор и Фрэнсис, стр. 124–130.
- Этаярадж, К. и Джурафски, Д. 2020. «Полезность в глазах пользователя: критика таблиц лидеров НЛП», *Препринт ArXiv ArXiv:2009.13888* .
- Этудо, У., Юн, В. и Лю, Д. 2017. «Финансовый концептуальный картограф элементов (FinSEM) для взаимодействия с XBRL: использование метода M3 Plus», *Системы поддержки принятия решений* (98), Elsevier, стр. 36–48.
- Фань, В., Лю, С., Лу, П. и Тянь, К. 2018. *Выявление числовых несоответствий в графиках* , представлено в Трудах Международной конференции по управлению данными 2018 года, стр. 381–393.
- Галисон, П. Л. и Д'Агостино, С. 1987. *Как заканчиваются эксперименты* , Чикаго, Иллинойс: Издательство Чикагского университета, Чикаго.
- Джермонпре, М., Ховорка, Д.С. и Коллопи, Ф. 2007. «Теория проектирования индивидуальных технологий», *Журнал Ассоциации информационных систем* (8:6), стр. 351–367.
- Гонсалес-Уэрта, Дж., Бубакер, А. и Мили, Х. 2017. «Подход к реинжинирингу бизнес-процессов для преобразования моделей BPMN в программные артефакты», в *E-Technologies* , стр. 170–184.
- Грегор, С. и Хевнер, А.Р. 2013. «Позиционирование и представление исследований в области науки о дизайне для максимального воздействия», *MIS Quarterly* (37:2), стр. 337–355.
- Грегор, С. и Джонс, Д. 2007. «Анатомия теории дизайна», *Журнал Ассоциации информационных систем* (8:5), стр. 312–335.
- Грегор, С., Крузе, Л.С. и Зайдель, С. 2020. «Анатомия принципа проектирования», *Журнал Ассоциации информационных систем* (21:6), стр. 1622–1652.
- Грегори, РВ и Мунтерманн, Дж. 2014. «Исследовательская заметка — Эвристическое теоретизирование: упреждающее создание теорий проектирования», *Исследования информационных систем* (25:3), INFORMS, стр. 639–653.
- Губа, Э.Г. и Линкольн, Ю.С. 1994. «Конкурирующие парадигмы в качественном исследовании», *Справочник по качественному исследованию* (2:163–194), Калифорния, Sage Publications, стр. 105.
- Гампорт, П.Дж. 2007. *Социология высшего образования: вклад и его контексты* , Вашингтон, округ Колумбия: Издательство Университета Джонса Хопкинса + ORM.
- Хевнер, А., Марч, С., Парк, Дж. и Рам, С. 2004. «Наука проектирования в исследованиях информационных систем», *MIS Quarterly* (28:1), стр. 75–105.
- Хевнер А., Парсонс Дж., Брендель А.Б., Лукьяненко Р., Тифенбек В., Трембле М.К. и фон Брокке Дж. 2024. «Прозрачность в исследованиях в области дизайна», *Системы поддержки принятия решений* (182:1), стр. 1–19.
- Хойнингген-Хюэн, П. 2013. *Систематичность: природа науки* , Оксфорд, Великобритания: Oxford University Press.
- Иивари, Дж. 2015. «Различие и противопоставление двух стратегий исследований в области науки о дизайне», *Европейский журнал информационных систем* (24:1), Springer, стр. 107–115.
- Иивари, Дж., Ротвит Перлт Хансен, М. и Хадж-Болоури, А. 2021. «Предложение по минимальной оценке повторного использования принципов проектирования», *Европейский журнал информационных систем* (30:3), Тейлор и Фрэнсис, стр. 286–303.
- Йоханнессон, П. и Пьеронс, Э. 2014. *Введение в науку дизайна* , Берлин / Гейдельберг: Springer.
- Коорнееф, Х., Верхаген, В. Дж. и Курран, Р. 2020. «Структура поддержки принятия решений и прототип для оценки диспетчеризации воздушных судов», *Системы поддержки принятия решений* (135), Elsevier, стр. 113338.
- Кюхлер, В. и Вайшнави, В. 2012. «Структура разработки теории в исследованиях в области дизайна: множественные перспективы», *Журнал Ассоциации информационных систем* (13:6), стр. 395–423.
- Ларсен, К. и Бонг, Ч. Х. 2016. «Инструмент для решения проблемы конструктивной идентичности в обзорах литературы и метаанализах», *MIS Quarterly* (40:3), стр. 1–23.
- Ларсен, К.Р. и Беккер, Д.С. 2020. *Автоматизированное машинное обучение*

- для бизнеса: введение в точную, простую и быструю аналитику», Нью-Йорк, штат Нью-Йорк: Оксфорд, Великобритания.
- Ларсен, К. Р., Ховорка, Д., Деннис, А. и Уэст, Дж. Д. 2019. «Понимание слона: дискурсивный подход к идентификации границ и построению корпуса для статей по обзору теории», *Журнал Ассоциации информационных систем* (20:7), стр. 15.
- Ли, Дж., Ларсен, К. Р. и Аббаси, А. 2020. «TheoryOp: структура проектирования и система для раскрытия поведенческих знаний посредством онтологического обучения», *MIS Quarterly*, стр. 1–55.
- Линкольн, YS, и Губа, EG 1985. *Naturalistic Inquiry (m. 75)*, Thousand Oaks, CA: SAGE Publications.
- Лоури, П. Б., Гаскин, Дж., Хамферис, С. Л., Муди, Г. Д., Галлетта, Д. Ф., Барлоу, Дж. Б. и Уилсон, Д. В. 2013. «Оценка качества журнала и корзины журналов старших ученых Ассоциации информационных систем с помощью библиометрических показателей: добавляют ли экспертные оценки журналов ценность?», *MIS Quarterly*, стр. 993–1012.
- Лукьяненко, Р. и Парсонс, Дж. 2020. «Неопределенность теории проектирования: что это такое, как ее можно уменьшить и почему утонул белый медведь?», *Журнал Ассоциации информационных систем* (21:5), стр. 1–30.
- Лукьяненко, Р., Парсонс, Дж., Вирсма, Й. и Маддах, М. 2019. «Ожидание неожиданного: влияние выбора дизайна сбора данных на качество краудсорсингового пользовательского контента», *MIS Quarterly* (43:2), стр. 634–647.
- Мандивалла, М. 2015. «Создание и обоснование теории проектирования», *Журнал Ассоциации информационных систем* (16:5), стр. 314.
- Марч, СТ и Смит, ГФ 1995. «Проектирование и естественнонаучные исследования в области информационных технологий», *Системы поддержки принятия решений* (15:4), стр. 251–266.
- Матадамас-Эрнандес Х., Роман-Алонсо Г., Рохас-Гонсалес Ф., Кастро-Гарсия М.А., Букерш А., Агилар-Корнехо М. и Кордеро-Санчес С. 2012. «Параллельное моделирование сетей пор с использованием многоядерных процессоров», *Транзакции IEEE на компьютерах* (63:6), IEEE, стр. 1513–1525.
- Максвелл, Дж. 1992. «Понимание и валидность в качественном исследовании», *Harvard Educational Review* (62:3), Harvard Education Publishing Group, стр. 279–301.
- Мет, Х., Мюллер, Б. и Мэдче, А. 2015. «Проектирование системы анализа требований», *Журнал Ассоциации информационных систем* (16:9), стр. 2.
- Moules, NJ, McCaffrey, J. и Field, J. 2015. *Проведение герменевтических исследований: от философии к практике*, Нью-Йорк, штат Нью-Йорк: Peter Lang Publishing.
- Национальные академии наук, электротехники и медицины. 2022. *Онтология в поведенческих науках: ускорение исследований и распространение знаний: дайджест-версия*, Вашингтон, округ Колумбия: Издательство национальных академий.
- Ньюэлл, А. 1975. «Учебник по системам понимания речи», *Распознавание речи*, стр. 4–54.
- Ньютон, П. и Шоу, С. 2014. *Валидность образовательной и психологической оценки*, Хобокен, Нью-Джерси: Sage.
- Нунамейкер, Дж. Ф., Чен, М. и Пурдин, Т. Д. 1991. «Разработка систем в исследованиях информационных систем», *Журнал систем управленческой информации* (7:3), стр. 89–106.
- Онвугебузи, А. Дж. и Лич, Н. Л. 2007. «Достоверность и качественное исследование: оксюморон?», *Качество и количество* (41:2), Springer, стр. 233–249.
- Пяакконен Т., Кежляйнен Дж., Кескустало Х., Азопарди Л., Максвелл Д. и Ярвелин К. 2017. «Проверка имитированного взаимодействия для поисковой оценки», *Information Retrieval Journal* (20), Springer, стр. 338–362.
- Падманабхан, Б., Фанг, Х., Саху, Н. и Бертон-Джонс, А. 2022. «Машинное обучение в исследованиях информационных систем», *MIS Quarterly* (46:1), стр. iii–xix.
- Пефферс, К., Туунанен, Т. и Ниехавес, Б. 2018. «Жанры исследований в области науки о дизайне: введение в специальный выпуск об образцах и критериях для применимых исследований в области науки о дизайне», *Европейский журнал информационных систем* (27:2), стр. 129–139.
- Пефферс, К., Туунанен, Т., Ротенбергер, М.А. и Чаттерджи, С. 2007. «Методология исследований в области науки о дизайне для исследования информационных систем», *Журнал систем управленческой информации* (24:3), стр. 45–77.
- Пирамуту, С. и Досс, Р. 2017. «О решениях на основе датчиков для одновременного

- присутствия нескольких RFID-меток», *Системы поддержки принятия решений* (95), Elsevier, стр. 102–109.
- Прат, Н., Комин-Ваттиау, И. и Акока, Дж. 2015. «Таксономия методов оценки артефактов информационных систем», *Журнал систем управленческой информации* (32:3), стр. 229–267.
- Рамакришнан, М., Грегор, С., Шреста, А. и Соар, Дж. 2023. «Принципы проектирования платформенно-ориентированных общих ресурсов знаний с пояснительной реализацией», *Журнал Ассоциации информационных систем* (24:5), стр. 1413–1438.
- Рефсгаард, Дж. К., Ван дер Слуйс, Дж. П., Браун, Дж. и Ван дер Кёр, П. 2006. «Структура для работы с неопределенностью, вызванной ошибкой структуры модели», *Достижения в области водных ресурсов* (29:11), Elsevier, стр. 1586–1597.
- Робертс, П. и Прист, Х. 2006. «Надежность и валидность исследований», *Nursing Standard* (20:44), Издательская компания Королевского колледжа медсестер (RCN), стр. 41–46.
- Роземанн, М. и Весси, И. 2008. «На пути к повышению релевантности исследований информационных систем для практики: роль проверок применимости», *MIS Quarterly* (32:1), стр. 1–22.
- Салмон, WC 1998. *Причинность и объяснение*, Oxford University Press.
- Седракян, Г., Поэлманс, С. и Сноек, М. 2017. «Оценка влияния быстрого прототипирования с обратной связью на понимание семантики параллельных диаграмм состояний UML начинающими разработчиками моделей», *Информационные и программные технологии* (82), Elsevier, стр. 159–172.
- Сейн, М., Хенфридссон, О., Пурао, С., Росси, М. и Линдгрэн, Р. 2011. «Исследования дизайна действий», *MIS Quarterly* (35:1), стр. 37.
- Шульц, К. С., Риггс, М. Л. и Коттке, Дж. Л. 1998. «Необходимость развития концепции валидности в промышленной и кадровой психологии: психометрические, юридические и новые вопросы», *Current Psychology* (17), Springer, стр. 265–286.
- Тейлор, CS 2013. *Валидность и валидация*, серия «Понимание статистики», Нью-Йорк, штат Нью-Йорк: Oxford University Press, США.
- Томас, МА, Ли, Й. и Ли, А.С. 2022. «Обобщение артефакта информационных систем», *Исследования информационных систем* (33:4), INFORMS, стр. 1452–1466.
- Тифенбек, В., Гётте, Л., Деген, К., Тасик, В., Флейш, Э., Лалив, Р. и Стааке, Т. 2016. «Преодоление смещения значимости: как обратная связь в реальном времени способствует сохранению ресурсов», *Наука управления* (64:3), стр. 1458–1476.
- Туунанен, Т., Винтер, Р. и Вом Брокке, Дж. 2024. «Преодоление сложности в научных исследованиях в области дизайна», *MIS Quarterly* (48:2), стр. 427–458.
- Умапати, К., Пурао, С. и Бартон, Р. Р. 2008. «Проектирование решений по интеграции предприятий: эффективно», *Европейский журнал информационных систем* (17:5), Springer, стр. 518–527.
- Валеча, Р., Шарман, Р., Рао, Х.Р. и Упадхайя, С. 2013. «Модель диспетчерско-опосредованной связи для систем реагирования на чрезвычайные ситуации», *ACM Transactions on Management Information Systems (TMIS)* (4:1), ACM New York, NY, USA, стр. 1–25.
- Венейбл, Дж., Прис-Хейе, Дж. и Баскервиль, Р. 2016. «FEDS: структура оценки в исследованиях в области дизайна науки», *Европейский журнал информационных систем* (25:1), стр. 77–89.
- Венкатеш, В., Моррис, М.Г., Дэвис, ГБ и Дэвис, Ф.Д. 2003. «Принятие информационных технологий пользователями: на пути к единому взгляду», *MIS Quarterly*, стр. 425–478.
- Вайнхардт К., Клокер С., Хинц О. и ван дер Аалст В.М. 2020. «Гражданская наука в исследованиях информационных систем», *Инженерия бизнеса и информационных систем* (62), Springer, стр. 273–277.
- Винтер, С., Беренте, Н., Хоуисон, Дж. и Батлер, Б. 2014. «За пределами организационного «контейнера»: концептуализация социотехнической работы 21-го века», *Информация и организация* (24:4), стр. 250–269.
- Зайцев, А. и Манкинен, С. 2022. «Разработка приложений финансового образования для развития: применение исследований в области разработки действий в сельской местности Камбоджи», *Европейский журнал информационных систем* (31:1), Тейлор и Фрэнсис, стр. 91–111.

БИОГРАФИИ

Кай Р. Ларсен [0000-0002-8812-9866] — профессор информационных систем в Школе бизнеса Лидса, Университете Колорадо в Боулдере. Он является приглашенным преподавателем на кафедре информационных наук, профессором Университета Агдера и научным консультантом Gallup. Кай наиболее известен тем, что предоставил практическое решение головоломки Джингла Эдварда Торндайка (1904) и своим вкладом в семантическую теорию ответов на опросы (STSR). Он получил премию *INFORMS Design Science Award* за 2019 год, премию *Herbert A. Simon Design Science Award* за 2020 год и премию *AIS Education Innovation Award* за 2023 год.

Роман Лукьяненко [0000-0001-8125-5918] — доцент Школы коммерции Макинтайра, Университета Вирджинии. Его научные интересы включают управление данными и методы исследования (валидность и искусственный интеллект). Роман активно разрабатывает идеи, инструменты и методы для улучшения управления данными и практики исследования. Эти решения получили крупные награды, включая премию *INFORMS Design Science Award*, золотую медаль генерал-губернатора Канады, премию *Herbert A. Simon Design Science Award*. Исследования Романа публиковались в *Nature*, *MIS Quarterly*, *Information Systems Research*, *ACM Computing Surveys*. Его статья 2019 года о качестве краудсорсинговых данных получила премию за лучшую статью в *MIS Quarterly*.

Роланд М. Мюллер [0000-0002-8706-7763] — профессор информационных систем в Берлинской школе экономики и права в Германии. Он получил докторскую степень по информационным системам в Свободном университете Берлина. Его исследовательские интересы включают онтологии и большие языковые модели для обзоров литературы, метатеоретический анализ, онтологическую инженерию, методы проектирования науки и инновации, управляемые пользователями в проектах ИИ. Он является автором трех книг и более 100 научных статей, а его работа получила множество наград, включая премию *Herbert A. Simon Design Science Award*. Помимо академической деятельности, Роланд внес вклад в инновации в качестве основателя и члена консультативного совета нескольких стартапов и имеет патент в области машинного обучения.

Веда С. Стори [0000-0002-8753-1553] — почетный профессор университета и профессор компьютерных информационных систем Tull, а также профессор компьютерных наук в бизнес-колледже Дж. Мака Робинсона, Университет штата Джорджия. Ее исследовательские интересы лежат в области управления данными, концептуального моделирования и науки дизайна. Она особенно интересуется оценкой влияния новых технологий на бизнес и общество с точки зрения управления данными. Доктор Стори является членом колледжа старших ученых AIS и руководящего комитета Международной конференции концептуального моделирования. Она является лауреатом премии Питера П. Чена Премия, ER, AIS и INFORMS.

Джеффри Парсонс [0000-0002-4819-2801] — профессор университетских исследований и профессор информационных систем на факультете делового администрирования Мемориального университета Ньюфаундленда в Канаде. Его исследовательские интересы сосредоточены на том, как лучше представлять человеческие концептуализации мира в данных. Его работа по этой и смежным темам была опубликована в нескольких дисциплинах. Исследования Джеффа были отмечены несколькими способами, включая статью года *MISQ* (2019), премию *AIS Senior Scholars Paper Award* (2020) и премию *INFORMS Design Science Award* (2014). Он является членом Ассоциации информационных систем, выдающимся научным сотрудником Технического университета Дрездена, старшим научным сотрудником Schoeller и научным сотрудником ER. Он работал старшим редактором в *MIS Quarterly* и в настоящее время является старшим редактором в *Information Systems Research*.

Дебра Вандер Мир [0000-0002-5930-6667] — профессор информационных систем в Международном университете Флориды (FIU). Ее текущие исследовательские интересы сосредоточены на применении концепций компьютерной науки и информационных систем к проблемам реального мира. Ее работы публикуются в ведущих журналах и на конференциях в этих областях. До прихода в FIU она десять лет проработала в промышленности, в том числе в стартапе, финансируемом венчурным капиталом. Она имеет степень бакалавра Джорджтаунского университета, степень магистра наук в области систем управления информационными системами Университета Аризоны и докторскую степень в области

компьютерных наук Технологического института Джорджии.

Дирк С. Ховорка [0000-0001-7049-5617] — профессор дисциплины «Бизнес-информационные системы» в Сиднейском университете, Австралия. Он получил докторскую степень по информационным системам в Университете Колорадо и имеет степень магистра по междисциплинарным телекоммуникациям и степень магистра по геологии. Его текущие исследования изучают, как научные и общественные практики порождают возможные «миры» посредством теории, проектирования и концептуализации будущего. Его исследования сосредоточены на *спекулятивных подходах* к «знанию» относительно технологий, общества и биофизических сред, признавая, что будущее

неумолимо сопротивляется эмпирическому знанию и что беспрецедентные социотехнические изменения делают наши знания о прошлом менее показательными для будущих состояний. Дирк — старший редактор JAIS (Research Perspectives), член редколлегии ISR и сопредседатель минитрека HICSS «Информирование исследований: взаимодействие с будущим» и трека ECIS «Будущее как место исследования». Он является почетным обладателем премии Beta Gamma Sigma Professor of the Year 2018, премии Wayne Lonergan Award за выдающееся преподавание (2018), премии Dean's Award за преподавание (2022). Дирк — соавтор лучшей статьи AIS 2011 года «Вторичный дизайн: случай исследования поведенческого дизайна».

Приложение А. Общее использование типов валидности в дизайн-науке

Мы изучили, в какой степени Design Science Validity Framework, переведенный в существующие 70 определений валидности, отсортированных в нашей структуре в правом столбце Таблицы 7, использовался в 199 научных статьях по дизайну, опубликованных в журнале AIS Senior Scholars' Basket of Eight в период с апреля 2004 года по декабрь 2017 года. Каждая статья была изучена на уровне предложений и сравнена с нашим списком типов валидности с использованием регулярных выражений на уровне слов. Первый автор прочитал все полученные предложения и исключил результаты поиска, где их регулярные выражения давали чрезмерные ложные срабатывания. Например, такие термины, как *точность*, *Полнота* и *отзыв* имеют определенные, но многозначные значения в исследованиях качества данных и машинного обучения. Остальные термины дали десятки тысяч предложений, поэтому представленные результаты являются консервативными оценками.

Мы определили количество раз, когда валидность, соответствующая единице в структуре (таблица 7), использовалась по крайней мере один раз в статье, организованной по году и типу валидности. Мы пришли к выводу, что в научных статьях по дизайну не используются те же термины валидности, которые используются в других дисциплинах, за исключением некоторых, используемых для описания метрик из матрицы путаницы. Названия валидности, связанные с характерными типами валидности, использовались редко. Даже в рамках часто используемой категории типов валидности эффективности большинство обсуждаемых терминов, связанных с валидностью, возникли из мер матрицы путаницы в машинном обучении. Исключения были в использовании *типов валидности характеристики метода*, следующих наиболее используемых типов валидности, встречающихся в литературе по науке дизайна.

Мы обнаружили отсутствие согласованности и отсутствие фактического использования терминов валидности в научных работах по дизайну, что подразумевает значительную возможность для улучшения коммуникации оценки за счет большей согласованности в языке вокруг явных утверждений о валидности. Как только исследователи примут на себя обязательства по общим нормам валидности, согласованность коммуникации и отчетности должна улучшиться.