

4.1.3 Verallgemeinbarkeit

Im Folgendem werden die Kontextfaktoren beschrieben, die zu dem Zweck der Verallgemeinbarkeit zugeordnet wurden. Des Weiteren wird untersucht, in wie weit Autoren und Autorinnen über die Verallgemeinbarkeit ihrer Annahmen diskutieren.

Vorrangig werden lediglich die 59 Publikationen untersucht, die mit einem ACM Award für *Distinguished Paper* oder *Distinguished Artefact* ausgezeichnet wurden und/oder einen Badge erhalten haben *ACM Artifacts Evaluated Reusable* oder *ACM Artifacts Available*. Es wird angenommen, dass diese Publikationen eine höhere Qualität aufweisen und mehr auf die Verallgemeinbarkeit eingegangen wird.

Wie bereits in Kapitel 3 beschrieben, wird angenommen, dass die Kontexte mit dem Zweck der Verallgemeinbarkeit an bestimmten Stellen in einer Publikation vorkommen. Daher wurden lediglich folgende Stellen berücksichtigt: *Threats of Validity*, *Conclusion*, *Limitations*, *Future Work* und *Discussion*.

Innerhalb 34 Publikationen wurden 35 Kontextfaktoren identifiziert. Davon beziehen sich 24 Kontextfaktoren auf die Verallgemeinbarkeit, 8 auf die Abgrenzung des Themas, zwei auf das Thema selbst und ein Kontextfaktor beschreibt Details der Publikation. Die Kontextfaktoren bezüglich der Verallgemeinbarkeit wurden in der Tabelle 4.1 zusammengefasst und in vier Faktorengruppen klassifiziert: Anwendung, Zielgruppe Programmiersprachen und Testtypen.

ID	Kontextfaktor	Klassifikation	Vorkommen
b535	Crowdtesting Plattform	Anwendung	TtV
b535	funktionale Tests	Testtypen	TtV
b535	Usability Tests	Testtypen	TtV
b535	Security Tests	Testtypen	TtV
b535	Performanz Tests	Testtypen	TtV
a435	open-source Systeme	Anwendung	TtV
a435	kommerzielle close-source Systeme	Anwendung	TtV
a949	convolutional neural network	Anwendung	TtV
b110	TensorFlow, Keras, Pytorch	Anwendung	TtV
b159	Philly und andere DL Plattformen	Anwendung	TtV
a691		Programmiersprachen	L
a309	Entwickler/innen, industrielle Teams	Zielgruppe	TtV
a481	große, industrielle Apps	Anwendung	C
a026	other domains with partially ordered constructs	Anwendung	C
a075	other applications and grammar	Anwendung	TtV
b459	SMT solver, other classes of solvers	Anwendung	C, FW
b086	Python, other languages, language-agnostic	Programmiersprachen	D, L
b061	specific scenarios	Anwendung	C
b447		Programmiersprachen	TtV
b447	real-world Java programs	Anwendung	C
a847	active analyzer and real world programs	Anwendung	C
a752	OSS Datensatz mit Puppet Skripten	Anwendung	TtV
a752	across languages	Programmiersprachen	TtV
a457	Android Apps	Anwendung	TtV
a666	Softwareentwicklungsfirmen, Arbeitsplatz	Zielgruppe	C
b183	große, open-source Softwareprojekte	Anwendung	TtV
a641	open-source Projekte, close-source Projekte	Anwendung	TtV
a210	Java und Python	Programmiersprachen	L, T
a714	control applications	Anwendung	TtV
a714	convolutional neural network	Anwendung	TtV

Table 4.1: Kontexte bezüglich der Verallgemeinbarkeit an den Stellen *Threats to Validity* (TtV), *Conclusion* (C), *Discussion* (D) *Limitations* (L), *Future Work* (FW)

In ungefähr 3/4 der Publikationen (72,8 %) ist die Sektion *Threats to Validity* vorhanden. In 41,8% dieser Publikationen wird diese Sektion in interne, externe und construct Validität aufgeteilt und in 20,9% in interne und externe Validität. Jedoch wird meistens mehr auf die interne Validität eingegangen und beschrieben welche Maßnahmen getroffen wurden, um diese zu stärken und um Fehler, Bias oder Effekte zu vermeiden. Die externe Validität wird meist nur knapp beschrieben und es werden sehr wenig Aussagen zur Verallgemeinbarkeit getätigt.

Es fällt auf, dass anstatt über die Verallgemeinbarkeit zu diskutieren, die Studie und das Vorgehen erneut beschrieben wird. In vielen Publikationen (39,5%) wird auf die Evaluation verwiesen, wie auf die Popularität der genutzten Benchmark ¹ und die Auswahl der evaluierten Testsubjekte. In 8 Publikationen (18,6 %) wird auf die Diversität oder auf die Limitierung der Teilnehmenden verwiesen. ²

	#
Aufteilung in internal, external und construct validity	18
Aufteilung in internal und external validity	9
Beschreibung wie die interne Validität bestärkt wird	7
Sektion: <i>Threats to Validity</i>	43
Keine Sektion <i>Threats to Validity</i> vorhanden	16
Limitations	15
Future Work	24
Kein Fokus auf Verallgemeinbarkeit	2
Keine Stellungnahme zur Verallgemeinbarkeit	11
Behauptung: erstmalige Studie	26
Verweis auf der Diversität / Limitierung der Teilnehmenden	8
Verweis auf andere Studien (scope, scale)	2
Verweis auf eigenen Scope, Scale	7
Verweis auf Vorgehen, Beschreibung der Studie	7
Verweis auf Evaluation	17
Verweis auf Implementierung	3

Table 4.2: Analyse der Diskussion zur Verallgemeinbarkeit

¹"Our use of this well-known and widely-used benchmark of real-world defects aims to ensure our results generalize" - *"Causal Testing: Understanding Defects' Root Causes"* von Brittany Johnson, Yuriy Brun und Alexandra Meliou [a087]

²"To bolster these observations we subsequently used a more diverse interview sample, which included participants from both large and small employers." - *"A Tale from the Trenches: Cognitive Biases and Software Development"* von Souti Chattopadhyay, Nicholas Nelson, Audrey Au, Natalia Morales, Christopher Sanchez, Rahul Pandita und Anita Sarma [a654]

In zwei Publikationen wird ausgesagt, dass der Fokus der Studie nicht auf die Verallgemeinbarkeit liegt³ und in 11 Publikationen (18,6%) wird ausgesagt, dass die Ergebnisse möglicherweise nicht generalisierbar sind.⁴ In diesen Aussagen wurden zwar Kontextfaktoren gefunden bezüglich der möglichen Verallgemeinbarkeit. Jedoch werden wenig Aussagen über diese Kontexte gemacht und außerdem wird nicht über die Verallgemeinbarkeit der Ergebnisse auf diese Kontexte argumentiert.

Des Weiteren fällt auf, dass in mehr als 1/4 der Publikation (27,1 %) die Sektion *Threats to Validity* nicht vorhanden ist. Da Kontextfaktoren auch in anderen Sektionen vorkommen können, wurde untersucht, ob neben der *Conclusion*, Sektionen vorhanden sind, um die Begrenzungen der Arbeit und die zukünftige Arbeit zu erläutern. In 10 der 16 der Arbeiten (62,5%) sind solche Sektionen oder Passagen vorhanden. Im Vergleich mit den Publikationen mit *Threats to Validity*, welcher in Tabelle 4.3 dargestellt ist, ist ein Unterschied in der Häufigkeit der Sektion *Limitations* bemerkbar. Es kann daher geschlussfolgert werden, dass in Publikationen ohne *Threats to Validity* mehr über die Begrenzungen der Arbeit diskutiert wird, statt über die Validität und die Verallgemeinbarkeit.

	#	<i>Limitations</i>	<i>Future Work</i>	L + F
<i>Threats to Validity</i> vorhanden	43	3 (6,9%)	12 (27,9%)	6 (13,95%)
<i>Threats to Validity</i> nicht vorhanden	16	4 (25%)	4 (25%)	2 (12,5%)
insgesamt	59	7 (11,8%)	16 (27,1%)	8 (13,5%)

Table 4.3: Analyse der Vorkommen von Sektionen zu *Threats to Validity*, *Limitations* (L) und *Future Work* (F)

Es ergibt sich eine Schnittmenge von 27 Publikationen (45,7%), in denen nicht oder nur oberflächlich auf die Verallgemeinbarkeit eingegangen wird, sodass Lücken in der Diskussion der Verallgemeinbarkeit bestehen. Lediglich in 25 Publikationen (42,3%) wurden Kontexte bezüglich der Verallgemeinbarkeit gefunden.

³"While desirable, generalizability was not the main focus of this study" - "A Tale from the Trenches: Cognitive Biases and Software Development" von Souti Chattopadhyay, Nicholas Nelson, Audrey Au, Natalia Morales, Christopher Sanchez, Rahul Pandita und Anita Sarma [a654]

⁴"Our findings might not be generalizable to other systems." - "Towards the Use of the Readily Available Tests from the Release Pipeline as Performance Tests. Are We There Yet?" von Zishuo Ding, Jinfu Chen und Weiyi Shang [b435]