“Best Practice” without Evidence – Agile Software Methodology as Example

Andrei Ranete

Inland Norway University of Applied Sciences
ARanete@protonmail.com

Abstract. Despite the essentiality of education, and the widely known unscientific nature of expert opinion, education in general appears to be based on expert opinion. The example analyzed herein is of Agile software methodology, which is deemed a best practice and therefore taught in most IT studies, in Norway and most probably internationally. This is despite that it appears to be a well known fact within its respective field that the Agile methodology lacks scientific justification. A tertiary analysis was conducted to test this well known fact and to serve as basis for exploring what should be considered sufficient evidence for inclusion within official education. The result of the tertiary study is that, indeed, the evidence for the Agile methodology is scarce at best. A method to avoid such mistake is suggested, which could be valuable to science in general. This method entails employing philosophers of science, epistemologists, to counteract potential expert biases and verify the curriculum before it is accepted in official education.

Keywords: Education, Curriculum Quality, Tertiary Analysis, Expert Opinion, Best Practice, Agile Software.

1 Introduction

Quality education is of course essential for humanity, and one of United Nations’ sustainable development goals ([UN], 2021). Critical thinking is the basis of science, so education, and is therefore at least recommended within official education. Expert opinion is considered the lowest form of evidence, on about equal level with experience reports or anecdotes (for example Hadorn et al., 1996; Herman, Raybould, 2014). Despite these facts, an idea like Agile software methodology [Agile] is taught within the majority of information technology [IT] studies in Norway and most probably internationally (for example Inayat et al., 2015). Understanding Agile is not necessary for understanding this text, so it is not presented herein also due to lacking space, but other sources are available (for example Dybå, Dingsøyr, 2008). Teaching Agile is a strange phenomenon as it appears to be well known that it lacks scientific basis, even though it is considered “best practice”.

This circumstantial evidence inspired a preliminary search to test its basis and 3 papers were identified. The first text discovered “a positive relationship between agile use and reported project success”, however with a low determination coefficient [R^2]

of 0,03 (Serrador, Pinto, 2015). Furthermore, its results were "significant" according to P testing and Wald testing, statistical tools known to cause type I errors, mistaken rejection of the null hypothesis (Bennet et al., 2009; Woolston, 2015; Liu et al., 2018). These details indicate very probable confounding factors. The second text concluded "(…) we find that the strength of the evidence in the current review regarding the benefits and limitations of agile methods, and for decisions related to their adoption, is *very low*" (Dybå, Dingsøyr, 2008). The third text discovered an increase and decrease in 2 Agile methods' success by approximately 1 point, with an error margin of approximately 1 point (Saltz et al., 2017).

The research question is twofold: What should be deemed good enough evidence for inclusion within official curriculum, for example considering Agile methodology? Agile was chosen because a concrete example seems more substantial, and Agile was specifically chosen more or less arbitrarily. Even though the rumor appears to be true, a systematic analysis about its evidence is deemed necessary to compellingly conclude that Agile does indeed not have much of a scientific basis.

2 Methodology of Agile Analysis

A tertiary analysis is deemed the most expedient way to address the secondary part of the research question. To avoid risks of personal bias, the selection and quality assessment of the secondary analyses is based on other relevant tertiary studies. Conducting this review follows standard procedure for systematic analyses: protocol development, identifying inclusion and exclusion criteria, searching studies, data extraction and synthesis. The exception is the exclusion of quality assessment for the tertiary reviews as a quaternary review is outside the scope herein. This chapter describes the protocol, following the methodology of one of the most influential articles in the field (Dybå, Dingsøyr, 2008), excepting the preventative approach mentioned initially to avoid personal bias. The inclusion criterion is being a tertiary study focusing only on studies about the Agile methodology. The main exclusion criterion is not presenting quality assessment. Another exclusion criterion, for the convenience of this analysis, is not being in English, or Norwegian. All known standard databases within the relevant field are chosen: ACM Digital Library, ScienceDirect, IEEE Xplore, Web of Science, SpringerLink, Scopus.

The aptest keywords are: agile AND software AND tertiary. The search was committed only to titles, abstract and keywords, if possible. 334 results were provided and, after assembly in a document, 301 unduplicated texts were found. All titles, with their first authors, that appeared to be potentially relevant, were noted in another document, including unspecific titles that could not be determined without further evaluation. 38 remained. After verifying the abstracts, 1 remained (Hoda et al., 2017).

This tertiary study included 28 secondary studies, so 27 are analyzed herein as one was inaccessible. Their main results are simply presented in a table. Main results should be presented in the results or discussion. Main results herein mean:

- any general statement about the efficiency of the Agile methodology, for example change in profitability or other measure of success, if existent, and

- self-reported effect or strength of evidence, if existent.

Clearer synthesis was impossible due to the reviews being diverse. Comparing the results otherwise may be impossible anyway as they are mostly presented in a narrative manner (for example Curcio et al., 2019; Cruzes, Dybå, 2010).

3 Results of Agile Analysis

Table 1. Tertiary analysis, quality score [Q#], criteria of CRD's DARE on a scale from 0 to 4

Text	Q #	Main Results
Global software engineering and agile practices: a systematic review, 2011, Jalali, S., Wohlin, C.	3	53 success stories were reported, but no mention of how successful, probably because most papers were opinion and experience reports. "In the majority of the study papers, the applied Agile method is addressed as "Agile", and distribution setting is mentioned as "distribution team" without any detailed information. It indicates the incompleteness of the contextual and background information in the current literature."
Reconciling software development models: a quasi-systematic review, 2012, Magdaleno, M., Werner, C. M. L., de Araujo, R. M.	3	This presents benefits and risks of combining different methodologies, including Agile, but does not comment specifically on Agile. Yet, it indicates a further lack of research in the field: "(…) in the technical literature, there is a scarcity of reports about reconciliation among all three software development models. By the end of the review, only a single paper (Theunissen et al., 2008) had been found covering all of them. Theunissen et al. (2008) discuss the challenges for reconciliation between agile and FOSS development models within a corporate culture."
Developers motivation in agile teams, 2012, Melo, C. de O., Santana, C., Kon, F.	3	This presents motivation in connection to Agile, so is not very relevant herein. Yet, this paragraph seems to be aptest for summarizing its findings: "Agile practices were not explicitly cited as a (de)motivator factor. Instead, they were related to many motivators, (…). Other practiceswere cited as support for motivators such as code sharing, sustainable pace and self-management."
An investigation into agile methods in embedded systems development, 2012, Albuquerque, O.,	3	Some "significant" benefits of Agile usage are mentioned, but not precisely elaborated. Main challenges and limitations are cultural, developers being unwilling to accept Agile.

Antonino, P. O., Nakagawa, E. Y.		How significant and reliable the results are, is not discussed.
Agile product line engineering-a systematic literature review, 2011, Díaz, J., Pérez, J., Alarcón, P., Garbajosa, J.	4	This focuses on a combination between Agile and something else, and so relevant information cannot be extracted. No strength of evidence is discussed.
A systematic mapping on agile UCD across the major agile and HCI conferences, 2015, da Silva, T. S., Silveira, F. F., Silveira, M. S., Hellmann, T., Maurer, F.	3	"The main benefit reported in the mapped studies is improved communication. The second benefit is improved usability. This benefit should be obvious; after all, this is the end goal of this integration. Other interesting benefits reported are the improved visibility and the improved business analysis." "As already mentioned, the majority of the studies did not provide the research method in the abstract. Case Study and Grounded Theory were the most common research methods in this field."
Factors limiting industrial adoption of test driven development: a systematic review, 2011, Causevic, A., Sundmark, D., Punnekkat, S.	3, 5	Test-Driven Development appears to be advantageous for the quality of code, but disadvantageous in 7 other aspects, like development time and design. Evidence is heterogenous, so difficult to judge. Plus, the strength of correlations is not mentioned, probably because it is difficult to judge.
Empirical studies on quality in agile practices: a systematic literature review, 2010, Sfetsos, P., Stamelos, I.	4	Test-Driven Development decreased defects by 0%-90%. Most other results are vaguer/more contradictory. Pair Programming increased design and code quality by 15%-65%, plus many other much less studied advantages and disadvantages. Non-specific improvements reported about Agile XP. "The generalizability of the results may be hindered also by the uncontrolled variables and metrics used in some of the studies. It is impossible to directly compare the results of these studies."
Using scrum in global software development: a systematic literature review, 2009, Hossain, E., Babar, M. A., Paik, H.	4	No relevant results can be extracted as only challenges of using scrum globally are discussed. "(…) only 4 studies (20%) included in this SLR are empirical studies and all of them are industrial case studies. Rest of the 16 studies (80%) are classified as "lesson learned" or industrial experience reports. Hence, we conclude that there is a little empirical evidence based report-

		ed on the use of Scrum practices in GSD context."
User-centered design and agile methods: a systematic review, 2011, da Silva, T. S., Martin, A., Maurer, F., Silveira, M.	3	"Regarding content, (…) results (need of an initiative, initiative proposal, lessons learned and recommendations)". "The primary work underlying this systematic review lacks sound controlled studies."
Systematic literature review on the characteristics of agile project management in the context of maturity models, 2014, Chagas, L. F., Carvalho, D. D., Lima, A. M., Reis, C. A. L.	2, 5	No relevant information can be extracted, especially since this focuses of a combination between Agile and something else. "Although many studies present practical experience in joint application, the majority refers to industry reports. Therefore, the area needs more rigorous studies using, for example, case studies and / or experimental studies."
The role of communication in agile systems development, 2013, Hummel, M., Rosenkranz, C., Holten, R.	3	"Positive impact: informal communication is reported as one of the key success factors of agile SD, especially in a distributed environment. Negative impact: informal communication is also reported to be challenging and overwhelming in case of many stakeholders or for maintaining knowledge." "(…) our review shows that our current state of knowledge on the precise role of communication in agile SD and its impact on SD success is limited because previous results are scattered, inconclusive, as well as contradictory."
Empirical studies of agile software development: a systematic review, 2008, Dybå, T., Dingsøyr, T.	4	Great increases in productivity for first iterations, but almost none in last iterations. The overall increase is questionable. Findings on product quality are inconclusive. "(…) we find that the strength of the evidence in the current review regarding the benefits and limitations of agile methods, and for decisions related to their adoption, is *very low*."
Approaches to agile adoption in large settings: a comparison of the results from a literature analysis and an industrial inventory, 2010, Rohunen, A., Rodriguez, P., Kuva-	3, 5	"(…) it was concluded that we could find no study which clearly and deeply indicates how to adopt agile methods." "Like in the literature analysis, the agile methods to be adopted were not specified in the majority of the industrial inventory research material, but in many cases Agile Scrum, XP and lean development were discussed. The findings of the industrial inventory are presented as gen-

ja, P., Krzanik, L., Markkula, J.		eralized results, independent of any specific agile methods."
Exploring principles of user-centered Agile software development, 2015, Brhel, M., Meth, H., Maedche, A., Werder, K.	4	"However, the translation of UCASD practices from academia into practice remains challenging, future empirically grounded research is required." "(…) the evaluation was based on the judgment and experience of the authors. Other scholars might have judged these articles differently. The same limitation applies to the coding of each paper using the presented coding system."
Using metrics in agile and lean software development, 2015, Kupiainen, E., Mäntylä, M. V., Itkonen, J.	3, 5	No relevant information can be extracted, as interestingly this focuses on metrics, but not the data collected with the metrics. "Deciding which Agile method was used in the cases was difficult. On the other hand, it is quite natural that cases use many aspects from multiple Agile methods."
Empirical studies of geographically distributed agile development communication challenges: a systematic review, 2016, Alzoubi, Y. I., Gill, A. Q., Al-Ani, A.	4	No relevant information can be extracted, as this does not mention the efficiency of Agile, most probably due to a lack of "confirmatory and explanatory studies." "It is still arguable whether agile practices can be effectively scaled up and used in GDAD environments due to communication challenges. Despite its acknowledged importance, we found that our knowledge about GDAD communication in practice is limited. This is the result of the study results being scattered, inconclusive, and ambiguous and scarcely any studies opening up the communication process or focusing on the social interaction and behaviour of teams as part of the research."
Towards an agile it organisation: a review of prior literature, 2008, Tapanainen, T., Hallanoro, M., Päivärinta, J., Salmela, H.	2	No Access
Effort estimation in agile software development: a systematic literature review, 2014, Usman, M., Mendes, E., Weidt, F., Britto, R.	3, 5	"These techniques used different accuracy metrics (…) to assess prediction accuracy, which in most cases did not turn out to meet the 25% threshold". "(…) we found very few studies reporting all the required elements appropriately e.g. 40% of the studies have not described the exact agile method used in the study, 24% of 88 them have not described the development activity that is subject to estimation, and 26% of the studies have

		not used any accuracy metric."
Agile methods tailoring – a systematic literature review, 2015, Campanelli, A. S., Parreiras, F. S.	2, 5	No relevant information can be extracted, as Agile efficiency is not mentioned. "Regarding to the research type used by the analyzed papers, solution proposal is the main research type adopted (42.8%) and experience papers is the second one (26.8%). This shows that the majority of the papers are either proposing new approaches for agile methods tailoring or reporting approaches already used in practice but not necessarily created based on scientific methods."
A systematic literature review for agile development processes and user centred design integration, 2014, Salah, D., Paige, R., Cairns, P.	4	"We did not find any rigorous controlled experiments. Therefore, it seems that the evidences in the suitability of agile methods and pros and cons of the methods need more research." "The results showed that the research is rather scattered and mainly driven by industry reports. It was found that there is no one method for the diverse world of embedded systems development, but many emphasize different viewpoints."
Agile methods for embedded systems development - a literature review and a mapping study, 2013, Kaisti, M., Rantala, V., Mujunen, T., Hyrynsalmi, S., Könnölä, K., Mäkilä, T., Lehtonen, T.	3	"It was found that there are embedded domain-specific problems about agile methods that need to be solved before agile methods can be successfully applied to the embedded domain." "To some extent, there are studies that address these issues, but the amount of evidence still remains scarce."
Agile testing: a systematic mapping across three conferences – understanding agile testing in the XP/Agile Universe, agile and XP conferences, 2013, Hellmann, T. D., Chokshi, A., Abad, Z. S. H., Pratte, S., Maurer, F.	3	No relevant information can be extracted, as Agile efficiency is not mentioned, this is a mapping study and as: "Unfortunately, it's again worth noting that abstracts of papers frequently did not include information crucial to understanding what the work was about."
Integrating agile and user-centered design. a systematic mapping	2, 5	"The majority of literature has found that a successful integration will be beneficial to the business and user. There are a number of factors and techniques that can be

and review of evaluation and validation studies of agile-UX, 2014, Jurca, G., Hellmann, T. D., Maurer, F.		considered to increase the chances of integration success." "Specifically, we note that there are very few papers reporting the results of focus group, action research, experiment, and survey studies." "The studies of our review have confirmed that there is a wide variety of how well the practices and artifacts which were described in literature can be applied."
A systematic literature review on agile requirements engineering practices and challenges, 2014, Inayat, I., Salim, S. S., Marczak, S., Daneva, M., Shamshirband, S.	4	"In our opinion, a larger number of empirical results are needed to provide evidence that agile RE practices resolve traditional RE challenges." "75% of the studies were exploratory in nature based on empirical investigations, newly proposed ideas and literature review papers." "The uneven distribution of authors across geographic regions means that the empirical evidence reported by the 21 studies could not be considered generalisable. (…) There are definitely differences in organisational culture, country-specific culture and social norms across organisations globally".
Using CMMI together with agile software development: a systematic review, 2015, Silva, F. S., Soares, F. S. F., Peres, A. L., de Azevedo, I. M., Vasconcelos, A. P. L. F., Kamei, F. K., de Lemos Meira, S. R.	4	"However, agile methodologies alone, according to the studies, were not sufficient to obtain the level desired, it being necessary to resort to additional practices." "(…) it is considered that the strength of evidence found is *low*, indicating that further research is very likely to have an important impact on confidence in the estimation of effects and is likely to change the estimate."
Information visualization for agile software development teams, 2014, Paredes, J., Anslow, C., Maurer, F.	3, 5	"The literature suggests there is a wide variety of novel software visualization tools to support practices such as code reviews, code exploration, and software maintenance." "Without clear supporting evidence that visualization tools can help improve development practices, developers won't make an effort to adopt these tools especially if it means they have to learn a new tool, learn how to interact with new devices, obtain expensive hardware, and maybe even change their working practices."
Factors affecting distributed agile	2,	"In the analysis of the literature we identify that most of the reported factors that affect distributed agile projects

projects: a systematic review, 2013, Matalonga, S., Solari, M., Matturro, G.	5	are more related to matters of distribution than the agile value proposition. We have also identified some successful case studies that have applied agile practices in the context of distributed software development." "Nevertheless, these are initial results and more research is needed to better generalize and build on their success."

4 Discussion

4.1 Agile Example

The results (Table 1) about the Agile methodology's efficiency have been inconsistent, ranging from a little detrimental to highly positive. In contrast, the Agile methodology's strength of evidence has been very consistent: Among all 27 secondary reviews, the ones that mention Agile's strength of evidence state it is low or very low. The secondary reviews themselves have an average Q# of approximately 3,4, meaning their strength of evidence is medium to high. The Agile methodology is based entirely or almost entirely on expert opinion and experience reports. In other words, all Agile evidence is preliminary, there is a clear lack of generalizable and/or controlled proof.

Limitations to this analysis exist of course. Perhaps the clearest limitation is the use of the term "Agile" as it is rather vague. However, this limitation is intrinsic as most research within the field appears to not specify the studied type of Agile (for example Jalali, Wohlin, 2011; Rohunen et al., 2010; Kupiainen et al., 2015). Most importantly, this tertiary analysis is hardly comprehensive. Without a comprehensive analysis, an utter lack of evidence cannot be ascertained. Moreover, even though my personal bias was prevented for study selection and quality assessment as they were based on other researchers, their potential biases may exist, however they seemed to have conducted sufficient preventative techniques (Hoda et al., 2017). Still, my personal bias as sole author may be considered reason for the extracted main results being so negative. The results are however consistent with the preliminary search and with other tertiary studies which discovered similar insufficiencies (Nurdiani et al., 2016; Curcio et al., 2019). Further considering the diversity and consistency of the selected secondary analyses, it is judged highly improbable that conclusive evidence exists on the subject. Indeed, if it truly is a known fact within the respective field that Agile lacks scientific basis, then this tertiary analysis might have been redundant.

Another aspect that may be considered a limitation in the current context is that the selected tertiary study focused exclusively on industrial studies. However, this may be considered a benefit, as industrial cases would probably be considered more relevant than ones in academic settings. Nevertheless, evaluating some academic instances should be useful, and fortunately a few such studies have been noted through this analysis. Academic studies on Agile seem consistent with the industrial ones. Particularly the Scrum Agile method appears negative to performance and to require most adaption, both industrially and academically (Saltz et al., 2017; Lous et al., 2017;

Masood et al., 2018). There are also at least 2 authors who do not believe eXtreme programming [XP], another form of Agile, is apt for tertiary education (Schneider, Johnson, 2005; Masood et al., 2018). These facts are particularly unpromising as XP and Scrum appear to be the most common, so most studied, Agile practices (Hoda et al., 2017).

Ultimately, despite potential appearances, it should probably be noted that I agree with Agile from a philosophical perspective. I believe indeed that all endeavors should be open-minded and judicious, as in agile, rather than strict and conventional. Yet, it is hopefully clear that being agile does not necessarily equal following a so-called Agile methodology. Agile is more of a philosophy rather than a methodology, and mindlessly following any methodology, regardless of whether it is called Agile, is the opposite of agile.

4.2 Curricular Quality Assessment

How is it possible that such unfounded idea was accepted within official education? It is acknowledged that social sciences have lower scientific standards, yet Agile clearly fulfils no standards. Agile specifically is clearly accepted because it is liked. It certainly sounds agile, and it appears that most people enjoy it (Dybå, Dingsøyr, 2008; Melo et al., 2012; Kropp et al., 2016; Kropp et al., 2018). Agile supporters are sometimes described as "evangelical" or almost "evangelical" in the literature ([BCS], 2004; Wray, 2010; Spurrier, Topi, 2017). I suppose the fundamental reason is that Agile seems intuitively best, however, the scientific method was created exactly because human intuition tends to be wrong, exactly because evangelism was realized not to be sufficient proof. Best practice is not necessarily good, despite seeming best.

The more important reason for how it is possible for such unfounded idea to be accepted, is a clear conflict of interests. Not only that Agile academics like it, more importantly their careers may be dependent on it. Without Agile, what else would they teach? How else would they create a baseless and congenial course and then be paid teaching it? How the curriculum is decided, appears to be a surprizingly opaque process. I have not been able to find clear information on how curricula are decided internationally, something that merits future study, however at least in Norway, after direct communication with a representative of the Norwegian Department of Education and a representative of the University of Agder, the curriculum appears to be entirely based on expert opinion. In defense of expert opinion, there exist situations in which it is a necessary evil due to urgency (Tonelli, 1999). Perhaps the most common example, in medical science, expert opinion must ethically be used to treat people, as one cannot wait for every aspect of medicine to be comprehensively studied and of course the scientific method has its limitations (Tonelli, 1999). However, there is no real urgency in an educational context, as students' lives do not depend on learning for the sake of learning. Indeed, students would surely appreciate a more focused curriculum, not being forced to learn unfounded ideas. It is very interesting that expert opinion is universally considered insufficient evidence within science, except for official curriculum, the most essential component of education, science.

The reason why no better procedure is conducted may be that this was assumed to be a necessary evil. After all, if a field's respective experts are not aptest for deciding,

then who? After contemplation, the answer becomes clear: philosophers of science. Philosophers of science, more precisely epistemologists, are supposed to be experts of evidence, so they should be able to neutrally approve the substance of courses. Inspired by governmental separation of power (de Secondat, 1750), the curriculum should be decided as such: Respective experts can propose ideas, but only neutral philosophers can approve them, and only then experts can officially enforce them, teach them. Conflicts of interest are known problems within science, and this is a method to eliminate or at least minimize them.

Exactly how this should be conducted, is outside the scope herein, partially due to lack of space. The objective herein was to explain the problem with an example and then propose a solution. Nevertheless, epistemologists and institutional experts may be more qualified to explain how this could work anyway. Still, there is another essential aspect that should be noted: quality criteria. Even if philosophers should be neutral, what is the minimal requirement to be officially accepted? Fundamentally, this appears to be a case of the sand dune paradox. If you removed or added bits of sand one by one, at what point would the dune become or cease to be a dune? I do not think it is just an accurate remark of the imprecision of human language, for after all what exactly is a dune, but also an accurate remark on the perhaps inherent imprecision of human endeavors. It seems such decision must be mostly or entirely arbitrary as it is paradoxical. At least moderate strength of evidence is suggested herein as minimal requirement for approval within official education. This can be found through a tertiary analysis like this one, for example. Of course, this does not mean that pedagogues are not allowed to talk about anything without at least moderate strength of evidence, just that students must not be evaluated so be forced to learn illogical and/or unproven ideas.

5 Conclusion

Even though Agile might have been a weirdly specific example, it would be naïve to assume that it is the only example of inept course material. Agile is simply a perfect case about the objectionable pervasiveness of expert opinion. I can think of several more examples, which I would rather not discuss without proper context, and surely others can as well. Even though systematic analyses and science in general are time-consuming pursuits, they are obviously worth the future of humanity, quality education.

References

Albuquerque, O., Antonino, P. O., Nakagawa, E. Y. (2012). An investigation into agile methods in embedded systems development. *Computational Science and Its Applications – ICCSA 2012*, 576-591. https://doi.org/10.1007/978-3-642-31137-6_44

Alzoubi, Y. I., Gill, A. Q., Al-Ani, A. (2016). Empirical studies of geographically distributed agile development communication challenges: a systematic review. *Information & Management*, 53(1), 22-37. https://doi.org/10.1016/j.im.2015.08.003

Bennett, C. M., Miller, M. B., Wolford, G. L. (2009). Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: an argument for multiple comparisons correction. *NeuroImage*, 47(1), S39-S41. https://doi.org/10.1016/S1053-8119(09)71202-9

Brhel, M., Meth, H., Maedche, A., Werder, K. (2015). Exploring principles of user-centered Agile software development. *Information and Software Technology*, 61, 163-181. https://doi.org/10.1016/j.infsof.2015.01.004

British Computer Society (BCS). (2004). Agile evolution. *The Computer Bulletin*, 46(6), 18-19. https://doi.org/10.1093/combul/46.6.18

Campanelli, A. S., Parreiras, F. S. (2015). Agile methods tailoring – a systematic literature review. *Journal of Systems and Software*, 110, 85-100. https://doi.org/10.1016/j.jss.2015.08.035

Causevic, A., Sundmark, D., Punnekkat, S. (2011). Factors limiting industrial adoption of test driven development: a systematic review. *2011 Fourth IEEE International Conference on Software Testing, Verification and Validation*. https://doi.org/10.1109/icst.2011.19

Chagas, L. F., Carvalho, D. D., Lima, A. M., Reis, C. A. L. (2014). Systematic literature review on the characteristics of agile project management in the context of maturity models. *Communications in Computer and Information Science*, 177-189. https://doi.org/10.1007/978-3-319-13036-1_16

Cruzes, D. S., Dybå, T. (2010). Synthesizing evidence in software engineering research. In Proceedings of the 2010 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM '10). *Association for Computing Machinery*, 1, 1-10. https://doi.org/10.1145/1852786.1852788

Curcio, K., Santana, R., Reinehr, S., Malucelli, A. (2019). Usability in agile software development: A tertiary study. *Computer Standards & Interfaces*, 64, 61-77. https://doi.org/10.1016/j.csi.2018.12.003

Díaz, J., Pérez, J., Alarcón, P., Garbajosa, J. (2011). Agile product line engineering-a systematic literature review. *Software: Practice and Experience*, 41(8), 921-941. https://doi.org/10.1002/spe.1087

Dybå, T., Dingsøyr, T. (2008). Empirical studies of agile software development: A systematic review. *Information and Software Technology*, 50(9-10), 833-859. https://doi.org/10.1016/j.infsof.2008.01.006

Hadorn, D. C., Baker, D., Hodges, J. S., Hicks, N. (1996). Rating the Quality of Evidence for Clinical Practice Guidelines. *Journal of Clinical Epidemiology*, 49(7), 749-754. https://doi.org/10.1016/0895-4356(96)00019-4

Hellmann, T. D., Chokshi, A., Abad, Z. S. H., Pratte, S., Maurer, F. (2013). Agile testing: a systematic mapping across three conferences – understanding agile testing in the XP/Agile Universe, agile and XP conferences. *2013 Agile Conference*. https://doi.org/10.1109/agile.2013.10

Herman, R. A., Raybould, A. (2014). Expert opinion vs. empirical evidence. *GM Crops & Food*, 5(1), 8-10. https://doi.org/10.4161/gmcr.28331

Hoda, R., Salleh, N., Grundy, J., Tee, H. M. (2017). Systematic literature reviews in agile software development: A tertiary study. *Information and Software Technology*, 85, 60-70. https://doi.org/10.1016/j.infsof.2017.01.007

Hossain, E., Babar, M. A., Paik, H. (2009). Using scrum in global software development: a systematic literature review. *2009 Fourth IEEE International Conference on Global Software Engineering.* https://doi.org/10.1109/icgse.2009.25

Hummel, M., Rosenkranz, C., Holten, R. (2013). The role of communication in agile systems development. *Business & Information Systems Engineering*, 5(5), 343-355. https://doi.org/10.1007/s12599-013-0282-4

Inayat, I., Salim, S. S., Marczak, S., Daneva, M., Shamshirband, S. (2015). A systematic literature review on agile requirements engineering practices and challenges. *Computers in Human Behavior*, 51(B), 915-929. https://doi.org/10.1016/j.chb.2014.10.046

Jalali, S., Wohlin, C. (2011). Global software engineering and agile practices: a systematic review. *Journal of Software: Evolution and Process*, 24(6), 643-659. https://doi.org/10.1002/smr.561

Jurca, G., Hellmann, T. D., Maurer, F. (2014). Integrating agile and user-centered design. a systematic mapping and review of evaluation and validation studies of agile-UX. *2014 Agile Conference*. https://doi.org/10.1109/agile.2014.17

Kaisti, M., Rantala, V., Mujunen, T., Hyrynsalmi, S., Könnölä, K., Mäkilä, T., Lehtonen, T. (2013). Agile methods for embedded systems development - a literature review and a mapping study. *EURASIP Journal on Embedded Systems*, 2013(1). https://doi.org/10.1186/1687-3963-2013-15

Kropp, M., Meier, A., Biddle, R. (2016). Agile Practices, Collaboration and Experience. In P. Abrahamsson, A. Jedlitschka, A. D. Nguyen, M. Felderer, S. Amasaki, T. Mikkonen (eds.), *Product-Focused Software Process Improvement*, 416-431. https://doi.org/10.1007/978-3-319-49094-6_28

Kropp, M., Meier, A., Anslow, C., Biddle, R. (2018). Satisfaction, Practices, and Influences in Agile Software Development. *Proceedings of the 22nd International Conference on Evaluation and Assessment in Software Engineering 2018*. https://doi.org/10.1145/3210459.3210470

Kupiainen, E., Mäntylä, M. V., Itkonen, J. (2015). Using metrics in Agile and Lean Software Development - A systematic literature review of industrial studies. *Information and Software Technology*, 62(C), 143-163. https://doi.org/10.1016/j.infsof.2015.02.005

Liu, Y., Andersson, B., Xin, T., Zhang, H., Wang, L. (2018). Improved Wald Statistics for Item-Level Model Comparison in Diagnostic Classification Models. *Applied Psychological Measurement*, 43(5), 402-414. https://doi.org/10.1177%2F0146621618798664

Lous, P., Kuhrmann, M., Tell, P. (2017). Is Scrum Fit for Global Software Engineering? *IEEE 12th International Conference on Global Software Engineering [ICGSE]*, 1-10. https://doi.org/10.1109/icgse.2017.13

Magdaleno, M., Werner, C. M. L., de Araujo, R. M. (2012). Reconciling software development models: a quasi-systematic review. *Journal of Systems and Software*, 85(2), 351-369. https://doi.org/10.1016/j.jss.2011.08.028

Masood, Z., Hoda, R., Blincoe, K. (2018). Adapting agile practices in university contexts. *Journal of Systems and Software*, 144, 501-510. https://doi.org/10.1016/j.jss.2018.07.011

Matalonga, S., Solari, M., Matturro, G. (2013). Factors affecting distributed agile projects: a systematic review. *International Journal of Software Engineering and Knowledge Engineering*, 23(9), 1289-1301. https://doi.org/10.1142/s021819401350040x

Melo, C., Santana, C. A., Fon, F. (2012). Developers Motivation in Agile Teams. *38th Euromicro Conference on Software Engineering and Advanced Applications*. https://doi.org/10.1109/seaa.2012.45

Nurdiani, I., Börstler, J., Fricker, S. A. (2016). The impacts of agile and lean practices on project constraints: A tertiary study. *Journal of Systems and Software*, 119, 162-183. https://doi.org/10.1016/j.jss.2016.06.043

Paredes, J., Anslow, C., Maurer, F. (2014). Information visualization for agile software development teams. *2014 Second IEEE Working Conference on Software Visualization*. https://doi.org/10.1109/vissoft.2014.32

Rohunen, A., Rodriguez, P., Kuvaja, P., Krzanik, L., Markkula, J. (2010). Approaches to Agile Adoption in Large Settings: A Comparison of the Results from a Literature Analysis and an Industrial Inventory. In M. Ali Babar, M. Vierimaa, M. Oivo (eds.), Product-

Focused Software Process Improvement. PROFES 2010. *Lecture Notes in Computer Science*, vol 6156. https://doi.org/10.1007/978-3-642-13792-1_8

Salah, D., Paige, R., Cairns, P. (2014). A systematic literature review for agile development processes and user centred design integration. *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering - EASE '14*. https://doi.org/10.1145/2601248.2601276

Saltz, J., Ivan, S., Kevin, C. (2017). Comparing Data Science Project Management Methodologies via a Controlled Experiment. *Proceedings of the 50th Hawaii International Conference on System Sciences*. https://doi.org/10.24251/hicss.2017.120

Schneider, J.-G., Johnston, L. (2015). eXtreme programming: helpful or harmful in educating undergraduates? *Journal of Systems and Software*, 74(2), 121-132. https://doi.org/10.1016/j.jss.2003.09.025

de Secondat, C.-L. (1750). *The Spirit of Laws*. France.

Serrador, P., Pinto, J. K. (2015). Does Agile work? — A quantitative analysis of agile project success. *International Journal of Project Management*, 33(5), 1040-1051. https://doi.org/10.1016/j.ijproman.2015.01.006

Sfetsos, P., Stamelos, I. (2010). Empirical studies on quality in agile practices: a systematic literature review. *2010 Seventh International Conference on the Quality of Information and Communications Technology*. https://doi.org/10.1109/quatic.2010.17

Silva, F. S., Soares, F. S. F., Peres, A. L., de Azevedo, I. M., Vasconcelos, A. P. L. F., Kamei, F. K., de Lemos Meira, S. R. (2015). Using CMMI together with agile software development: a systematic review. *Information and Software Technology*, 58, 20-43. https://doi.org/10.1016/j.infsof.2014.09.012

da Silva, T. S., Martin, A., Maurer, F., Silveira, M. (2011). User-centered design and agile methods: a systematic review. *2011 AGILE Conference*. https://doi.org/10.1109/agile.2011.24

da Silva, T. S., Silveira, F. F., Silveira, M. S., Hellmann, T., Maurer, F. (2015). A systematic mapping on agile UCD across the major agile and HCI conferences. *Computational Science and Its Applications -- ICCSA 2015*, 86-100. https://doi.org/10.1007/978-3-319-21413-9_7

Spurrier, G., Topi, H. (2017). When is agile appropriate for enterprise software development? *Proceedings of the 25th European Conference on Information Systems [ECIS 2017]*, 2536-2545. Research in Progress: https://aisel.aisnet.org/ecis2017_rip/6

Tonelli, M. R. (1999). In defense of expert opinion. *Academic Medicine*, 74(11), 1187-1192. https://doi.org/10.1097/00001888-199911000-00010

United Nations (UN). (2021). Take Action for the Sustainable Development Goals. https://www.un.org/sustainabledevelopment/sustainable-development-goals/

Usman, M., Mendes, E., Weidt, F., Britto, R. (2014). Effort estimation in agile software development: a systematic literature review. *Proceedings of the 10th International Conference on Predictive Models in Software Engineering*. https://doi.org/10.1145/2639490.2639503

Woolston, C. (2015). Psychology journal bans *P* values. *Nature*, 519(7541), 1-9. https://doi.org/10.1038/519009f

Wray, S. (2010). How Pair Programming Really Works. *IEEE Software*, 27(1), 50-55. https://doi.org/10.1109/MS.2009.199