Empirical evaluation of reparability scoring systems for validity and reliability

Dangal, Sagar; Sandez, Sonia; Bolaños Arriola, Julieta; Faludi, Jeremy; Balkenende, Ruud

DOI
10.1016/j.resconrec.2025.108211

Publication date
2025

Document Version
Final published version

Published in
Resources, Conservation and Recycling

Citation (APA)
Dangal, S., Sandez, S., Bolaños Arriola, J., Faludi, J., & Balkenende, R. (2025). Empirical evaluation of reparability scoring systems for validity and reliability. *Resources, Conservation and Recycling*, *218*, Article 108211. https://doi.org/10.1016/j.resconrec.2025.108211

Resources, Conservation & Recycling 218 (2025) 108211

Contents lists available at ScienceDirect

Resources, Conservation & Recycling

journal homepage: www.sciencedirect.com/journal/resources-conservation-and-recycling

Empirical evaluation of reparability scoring systems for validity and reliability

Sagar Dangal [a], Sonia Sandez [b], Julieta Bolaños Arriola [a], Jeremy Faludi [a], Ruud Balkenende [a,*]

[a] TU Delft, Building 32, Landbergstraat 15, 2628CE, Delft, the Netherlands
[b] Universitat Jaume I, Department of Mechanical Engineering & Construction, Ave. Sos Baynat s/n, 12071, Castellón, Spain

ARTICLE INFO

Keywords:
Repair Scoring Assessment
Ease of Disassembly
Repair Scorecard
Circular Economy
Repair
Disassembly

ABSTRACT

The validity and reliability of four prevalent reparability scoring systems has been investigated by comparing scores of ten smart phones and six vacuum cleaners versus empirically measured repair times, as well as comparing hypothetical ideal and problematic scenarios. Ease of disassembly methods was also assessed for five smart TVs, four washing machines and six vacuum cleaners. The scoring systems studied were the French Reparability Index (FRI), Joint Research Centre Scoring System (RSS/JRC), iFixit, and ONR19202. Overall scores of products across scoring systems were relatively well correlated, indicating a fair amount of overall reliability. However, the variability in scores for the best and worst case of the same product was often larger than the differences between products. Validity was good for products that are easily repairable, but scorecards often failed to score low when repair is infeasible or too expensive. Repair scores greatly depend on disassembly; since some scorecards count numbers of disassembly steps and other scorecards use proxy times, these two methods were compared against empirical disassembly times for five vacuum cleaners, five televisions, and four washing machines. The proxy time method was found to be highly accurate for all three product categories; the steps method was less so. It indicated the relative ease of disassembly well for washing machines, but not for televisions or vacuum cleaners. Finally, this study proposes improvements to scoring methods, including a limiting factor approach and the development of clearer protocols, to ensure the scoring systems are robust, reliable, and can effectively guide sustainable product design.

1. Introduction

The short average lifespan of many products has significant environmental impacts. The European Commission's circular economy action plan aims to address this issue by encouraging the production of more durable and repairable products, establishing specific requirements for the durability and reparability of products (European Commission, 2020). The right to repair movement aligns with these objectives, promoting an increase in the ability of consumers to repair their own devices, and challenging manufacturer-imposed restrictions (Marikyan and Papagiannidis, 2024).

Reparability scoring systems have emerged as a crucial tool in achieving these objectives. They provide a semi-quantifiable measure of a product's reparability, serving as a valuable resource for policy makers, designers, and manufacturers looking to enhance product reparability. Moreover, these systems empower consumers by enabling them to make informed decisions when purchasing products (Marikyan and Papagiannidis, 2024).

Within this context, various scoring systems (or reparability indicators) have been developed (Hervier Marie and De, 2018; Bracquené et al., 2018; Cordella et al., 2019; European Commission, 2022; ONR 192102, 2014; Flipsen et al., 2019) to evaluate the reparability of products. These systems evaluate a variety of factors, including the ease of disassembly, the availability of information and spare parts, pricing, and software aspects. Four scoring systems, namely French Reparability index (FRI) (2024), Joint Research Centre scoring system (i.e. RSS or JRC scoring system) (Bracquené et al., 2018), iFixit (Flipsen et al., 2019), and ONR19202 (ONR 192102, 2014), are particularly noteworthy. They have either been implemented or have a high likelihood of implementation in their respective contexts. For instance, the FRI system is currently in use in France and could possibly be implemented in other European countries (UNEP Circularity Platform, 2024). Similarly, the

* Corresponding author.
E-mail address: a.r.balkenende@tudelft.nl (R. Balkenende).

https://doi.org/10.1016/j.resconrec.2025.108211
Received 8 July 2024; Received in revised form 14 February 2025; Accepted 20 February 2025
Available online 12 March 2025
0921-3449/

ONR19202 system has been established as the standard in Austria. The JRC scoring system is currently being adopted alongside an energy level scoring system across Europe (C et al., 2022). The iFixit scoring system has potential for implementation on its website. Moreover, these systems can provide product-specific scores and their usage instructions are publicly accessible.

For these scoring systems to be effective in policymaking and for assessment by consumer organizations, market surveillance authorities, and other stakeholders, they must be valid and reliable. Validity means that the scoring system accurately measures what it is intended to measure, while reliability means consistency and objectivity of the scores upon re-evaluation (Paper, 1993; Bannigan and Watson, 2009), avoiding variations caused by subjective interpretations. However, the current evaluation of validity and reliability of scoring systems is limited, and this paper intends to address this gap.

1.1. Validity and reliability of scoring systems

Previous studies have identified several validity issues in scoring systems. A paper by Barros et al. (2023) (Barros and Dimla, 2023) and a report by HOP (2020) (HOP, 2020) assessed the FRI scoring system and found that equal weighting of factors, such as the availability and cost of spare parts, as well as the removability of priority parts (Schischke et al., 2022) could mean that products which are impractical to repair could still receive a high score, as a poor score in one area might be compensated in other areas.

Several studies (Bracquené et al., 2018; Barros and Dimla, 2023; Bracquené et al., 2021) have compared scoring systems. These analyses help identify variances in reparability scores for identical products across different systems. If significant differences between scores cannot be justified, it may suggest the scoring system's validity needs further examination. These studies suggest that, within their contexts, scoring systems tend to provide similar scores. However, a comparison of the new iFixit, FRI, and JRC scoring system is currently absent. This can offer a comprehensive view of the similarities and differences between these scoring systems, providing insights on the extent of their validity.

Ease of disassembly is an important criterion in all reparability scoring systems (Barros and Dimla, 2023; Vanegas et al., 2018). However, the validity of methods for assessing ease of disassembly remains unclear. Current scoring systems employ two distinct methods: "proxy time methods", such as "eDiM" (Vanegas et al., 2018), and iFixit Proxy time (Flipsen et al., 2019) which estimate disassembly times assigning times to specific actions; respectively the "step method" (EN45554, 2020), which counts the number of disassembly steps required to disassemble a specific component. While proxy-time methods are suggested to be more representative of actual ease of disassembly than the step method (Bracquené et al., 2021), these have only been validated for a limited number of (ICT) products and there is a notable absence of comparing the results with the actual disassembly time for a larger range of products, like domestic appliances.

Reliability issues in the JRC and FRI scoring systems due to different interpretation of criteria have been identified in several papers and reports (HOP 2020, Boix et al., 2023). These include the way of addressing spare part bundles (i.e., multi-functional modules with multiple components brought together using non removable fasteners) during disassembly, provision of repair information, and fluctuations in spare part prices. However, the extent to which these issues influence scores, and their effect in other scoring systems is not clear.

Overall, a complete picture regarding validity and reliability of the current scoring systems is lacking. This paper aims to address this gap for the four scoring systems mentioned above. The overarching research question of this study is: How valid and reliable are the current scoring systems? This question is explored through the following activities and analyses.

- The scores of a variety of products in different product categories are compared across the different scoring systems, to examine the differences and their justification. While this comparison does not directly determine the validity of a scoring system, it provides indications of areas that may require further investigation in terms of validity.
- The validity of the scoring system is evaluated by checking how the scoring systems handles hypothetical scenarios where repair is considered not feasible.
- For ease of disassembly, the proxy time method and steps method are compared with measured disassembly times for a range of different products to test the validity of each method.
- The reliability of the scoring systems is determined first by scoring various products with best-case and worst-case interpretations per scoring criterion, then by identifying the cause and the extent of differences in scores.

In total, 16 products (10 smartphones and 6 vacuum cleaners) are evaluated from the perspective of self-repair and 3rd party professional repair. For assessing ease of disassembly 5 smart TVs are also investigated. Aspects related to OEM (authorized) repair are not assessed, due to lack of access to information and tools provided for authorized repair.

Based on the findings, the paper proposes recommendations to improve validity and reliability in the current scoring systems. These improvements may empower designers, repairers, consumer organizations, and policymakers to leverage these systems more effectively.

2. Scoring systems

This section provides an overview of the four scoring systems that will be analysed. The criteria applied in the scoring system, the products analysed in the study, and their weight per criteria is presented in Table 1.

2.1. Joint Research Centre (JRC) Scoring system

The JRC scoring system has been developed based on the preliminary draft of the standard EN45554, which concerns general methods for the assessment of the ability to repair, reuse, and upgrade energy-related products (EN45554, 2020) and the Benelux study on "reparability criteria for energy-related products" (Bracquené et al., 2018), and is intended to be implemented as part of the eco-design requirements for consumer products. This scoring system is mainly focused on professional repairers, with some criteria considered for consumer repair. It is currently suited for the assessment of vacuum cleaners, washing machines, smartphones, tablets, and TV's.

2.2. French Reparability Index (FRI)

The French Reparability Index (FRI) was published in 2020 and is also develop based on EN45554 standard (EN45554, 2020). French law requires companies to analyse and score their own products and publish the score. This reparability index would most likely be part of sustainability index in the future which would include reliability index. FRI is focused on professional repairers and consumers, and some criteria are also considered for producers. This scoring system is currently used to assess washing machine, vacuum cleaners, smartphones, laptops, tv, corded trimmers, lawn mower, pressure washers and dishwashers (Ministère de la Transition Écologique, 2024).

2.3. iFixit scoring system

This scoring system is mainly based on research publications by Flipsen et al. (2019, 2016) and input from researchers and repair experts. This scoring system has been developed as an objective alternative to the current scoring system used in the iFixit website (iFixit, 2021).

Table 1
Overview of FRI, JRC and iFixit scoring system. “PRO” = Producer, “CON” = consumer, “REP”= repairer, “M”=must have criteria.

Criteria	**FRI**					**JRC**					**Ifixit**			**ONR (192102)**		
	Sub criteria	**Smartphone**		**Vacuum cleaner**		**Sub criteria**	**Smartphone**		**Vacuum cleaner**		**Sub criteria**	**Smartphones**		**Sub criteria**	**Vacuum**	
		Weight	**Total criteria weight**	**Weight**	**Total criteria weight**		**Weight**	**Total criteria weight**	**Weight**	**Total criteria weight**		**Weight**	**Total criteria weight**		**Weight**	**Total criteria weight**
Ease of Disassembly	Disassembly step	10 %	**20 %**	10 %	**20 %**	Disassembly step	25 %	**55 %**	18 %	**55 %**	Disassembly time	21 %	**55 %**	Ability to disassemble	20 %	**33 %**
	Tools required	5 %		5 %		Tools required	15 %		18 %		Path of entry	21 %		Modularity	6 %	
	Fasteners type	5 %		5 %		Fasteners type	15 %		18 %		Tools required	13 %		Ease of de-soldering	8 %	
Informaiton availability	Type of information (REP, CON)	20 %	**35 %**	20 %	**40 %**	Type and cost of information (PRO, CON)	15 %	**15 %**	18 %	**18 %**	Availability of repair information	13 %	**17 %**	Availability of repair information (M)	38 %	**49 %**
	Information on update type	10 %		–		–	–		–		Visual cues	4 %		Telephone support (M)	–	
	Remote assistance availability (REP, CON)	5 %		20 %		–	–		–		–	–		Diagnosis design features (M)	11 %	
Spare part availability	Availability over time (PRO, RET, REP, CON)	15 %	**20 %**	15 %	**20 %**	Availability over time	–	**15 %**	9 %	**18 %**	Availabiltiy over time	5 %	**21 %**	Availabiltiy over time (M)	8 %	**11 %**
	Delivery time (PRO, RET, REP, CON)	5 %		5 %		Who is spare part available to	15 %		9 %		Who is spare part available to	17 %		Standarised parts (M)	3 %	
Spare part price	Ratio between part and product price	20 %	**20 %**	20 %	**20 %**	–	–		–		**–**	–	**–**	**–**	–	**–**
Software aspects	Software reset (PRO,REP, CON)	5 %	**5 %**	–	**–**	Availability over time	15 %	**15 %**	4,5 %	**9 %**	–	–	**–**	–	–	**–**
	–	–		–	**–**	Free avaiablity of update	–		4,5 %							
Health and safety	–	–		–		–	–	**–**	–	**–**	Tools risk	2 %	**4 %**	Warning signs	2 %	**4 %**
							–	**–**		**–**	Puncture risk	2 %		Information on series errors (M)	2 %	
Repair endorsement	–	–		–		–	–	**–**	–	**–**	Repair allowed by	4 %	**4 %**	–	–	**–**

This scoring system focuses on consumer repair and is currently only suited for smartphones.

2.4. ONR192102 (2014)

The ONR192102 scoring system (ONR 192102, 2014) has been developed through a collaboration between repairers and the Austrian Federal Ministry of Land, Forestry, Environment, and Water. It is intended for professional repairers. This system is suited to evaluate the durability and repairability of both brown and white goods.

3. Methods

This section presents the method used to score products with different scoring systems. These scores are then used to assess scoring systems ability to address a hypothetical scenario, to compare differences between scoring systems, and to assess best and worst-case interpretations per criterion. Also, the method used to analyse ease of disassembly is presented.

3.1. Scoring products

3.1.1. Products and priority parts for investigation

Vacuum cleaners and smartphones were evaluated since they can be scored with most (3/4) of the scoring system under investigation. Furthermore, these two products offer perspectives from two distinct product categories and design complexity: One being highly integrated and miniaturized and another being relatively large and mechanically intensive.

The products are scored based on the protocol presented in the scoring system. The FRI, JRC, and iFixit scoring systems incorporate the concept of 'priority parts' in their assessments. These parts are identified based on their likelihood of failure and functional relevance. However, what constitutes high and low failure likelihood, as well as functional relevance, can vary across different scoring systems. The priority parts for smartphones and vacuum cleaners, and its weight across various scoring systems is provided in supplementary material 9.1. For JRC and iFixit scoring system, the relative weight given to the priority parts is indicated. Priority parts for FRI are categorized in list 1 and 2. Ease of disassembly, tools required, and spare part price criteria are only applicable for list 2. The ONR system considers all parts in its assessment.

3.1.2. Product score determination

For scoring the products with each scoring system, the following protocols were used for disassembly assessment, and for obtaining information from the manufacturers and the internet.

3.1.2.1. Disassembly assessment protocol. The following protocol was used to score criteria related to disassembly.

1. The disassembly and reassembly process of the product to access all the priority parts is first determined.
   a. The official website was checked for a disassembly manual or instructions.
   b. If this is not available, other non-official disassembly/reassembly instructions websites such as iFixit and YouTube disassembly channels are consulted.
   c. If none of these sources provide the necessary information, the product is disassembled and reassembled in order to determine the disassembly process.
2. The product is tested to ensure that it functions as expected.
3. The device is then disassembled until each priority part is separated.
   a. Any non-standard accessories, such as display protectors or rubber bumpers, are removed at the start of the disassembly process. These accessories are not considered part of the disassembly.
   b. For battery-powered items, the shortest possible route to remove the battery is taken first as a safety measure.
   c. Where needed, adhesive was removed using, combination of heatgun, and isopropanol. All other operations were carried out using basic tools (Class A) (EN45554, 2020).
   d. If there is a risk of damaging components during further disassembly, we took steps to avoid this risk, even if it results in a longer disassembly path. For example, when replacing an iPhone 12 battery, the display was disconnected and removed to avoid tearing the display cables by movements when pulling on the pull tabs.
   e. The disassembly of a priority part ended when it reached permanent fixtures, such as soldering, welding, or thermal molding.
   f. If multiple tools are required during a single disassembly action, the tool is only changed once all fasteners requiring that specific tool have been removed.
4. The product is subsequently reassembled.
5. The product was tested again to confirm it works correctly and disassembly has not influenced its functionality.
6. The entire disassembly and reassembly process is recorded in top and side-view videos (see Supplementary material 9.2).
7. During the disassembly, the actions are described aloud, mentioning and describing the target component, its location, the tool being used, and the detached/reattached fasteners.
8. A disassembly map (De Fazio et al., 2021) for the product is made based on the video and audio recording.
9. Based on the disassembly map, the criteria related to disassembly and the disassembly sequence were extracted. The disassembly sequence considers the product fully assembled when analyzing each priority part's disassembly.

The 8th step in the disassembly assessment protocol evaluated the following criteria: a) disassembly steps; b) proxy disassembly time (Vanegas et al., 2018); c) fastener visibility; d) fastener type; e) path of entry; f) safety during repair; and g) tools required. The proxy time method used was based on data regarding the Disassembly and Reassembly Times (DaRT) obtained from monitoring and timing the disassembly and reassembly process of 12 washing machines (disassembly and reassembly) and 12 vacuum cleaners (disassembly and reassembly), 35 smartphones (deglueing and prying actions), and 7 tv (prying actions) (Ministère de la Transition Écologique, 2024). This method uses as a proxy the median time needed for a range of connections with a large number of specified tools. An overview of proxy time data for different tool-connection combinations can be found in Supplementary material 9.4. This method was preferred over eDiM, because the results are based on a larger variety of different products, don't need adaptation to tools used and the actions are simpler than the eDiM model.

3.1.2.2. Protocol to assess criteria that require information from manufacturers. To test the aspects related to the availability of repair information, the following protocol was used:

1. First, the local (EU, Netherlands (NL)) OEM website was searched to see if any repair guidance or links to repair guidance were available, including links provided by the OEM to YouTube or iFixit videos as well as service manuals.
2. A Google search was then conducted to identify if any service manuals, repair guides, or repair videos were available directly from the OEM. Note that unless suggested by the OEM directly, repair information made available by third parties (e.g., iFixit) is not considered.

3. Finally, the local customer service of the OEM was called and asked if they were able to provide the repair information necessary to repair the product.
   a. This was done first as a repairer and then as a consumer.
   b. If customer service directed us to another body, this was followed through (e.g., a technician or official repair partner).
   c. This was done until no further personnel could be contacted, and no further information could be retrieved by phone.

To test the aspects related to the price and availability of spare parts, the following protocol was followed.

1. First, the local (EU, NL) OEM website was searched to see if spare parts or links to authorized spare parts providers were available or suggested.
2. A Google search was then conducted to identify if any other web shops sold official spare parts. Non-official parts supplied by third parties (e.g., iFixit or another website) are not considered for the scoring but serve as a reference to determine if components are available as spare parts.
3. Finally, customer service was called and asked if they could provide or direct us to authorized priority parts purchasing.

The protocol for obtaining information was subsequently evaluated the following criteria: a) repair and diagnosis information; b) spare part availability; c) spare part price; d) policy on self-repair; e) information on updates; and f) remote assistance. The information (including the spare part price and availability) was checked during the analysis (November 2022–Jan 2023)

If multiple interpretations were possible in scoring a specific criterion, the scoring system protocol was rechecked to determine whether this was due to deviation from the protocol; if none was found, the different interpretations were considered as best- and worst-case scenario.

3.1.2.3. Researchers' co-agreement. Both the disassembly assessment and the scoring of the products were carried out by two researchers experienced in disassembly and product assessment. The results were checked against each other for co-agreement. Discrepancies between the results were first rechecked to determine whether they were the result of a deviation from the protocol. Remaining discrepancies were noted as difference in interpretation and provided input for evaluating reliability.

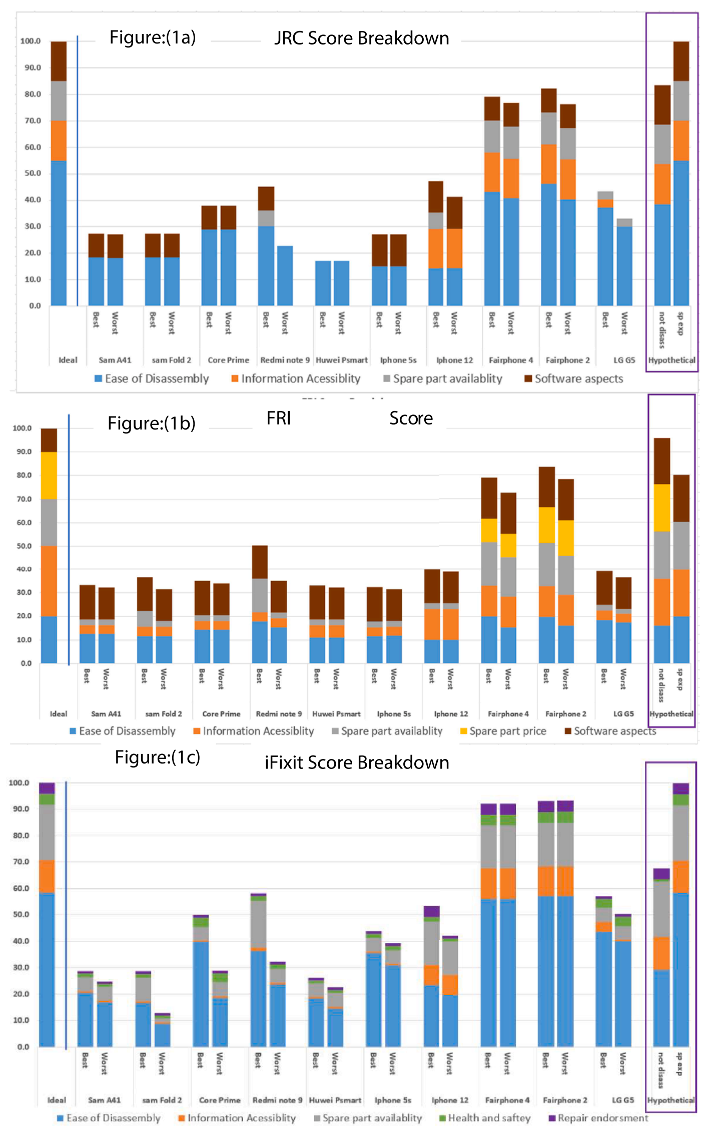

Fig. 1. Score breakdown for smartphones. 1a JRC score, 1b FRI score, 1c iFixit score, "sp exp " = expensive spare part; "not disass" = *a* priority part cannot be disassembled.

3.2. Analysis of scoring systems

3.2.1. Hypothetical scenarios

The validity of the scoring system was checked by creating hypothetical scenarios where repair is unfeasible and determining how different scoring systems handle the situation. In scenarios where it was clear that repair is not feasible, a valid scoring system should address this accordingly by giving it a low score. Therefore, two hypothetical scenarios were checked.

In the first scenario, a hypothetical product was assessed with a spare part price of one of its components (motor for a vacuum cleaner or screen for a smartphone) was over 50 % the price of a new product, while all other criteria of the scoring system were met in an ideal way.

Similarly, in the second scenario, a hypothetical product was assessed in which the priority part (motor for a vacuum cleaner or screen for a smartphone) is attached by a permanent fixture and cannot be disassembled, while again ideally meeting all other criteria.

Figs. 1 and 2 also show the 'ideal' scenario, which represents the maximum obtainable score, and illustrate the weight of the various scoring categories in the system under study.

3.2.2. Comparing different scoring systems

To investigate the relationship between different scoring systems and identify whether the scores agree on the rankings of better products vs worse products, a regression analysis was conducted.

For vacuum cleaners, Wilcoxon signed-rank test for scores on vacuum cleaners to determine whether the differences in scores were significant between the JRC and FRI scoring systems. Similarly, we conducted a Friedman's test for scores on smartphones to determine whether the differences in scores were significant between the three scoring systems (JRC, FRI, iFixit). These analysis methods were chosen since the scores were not normally distributed, dependent, and the number of samples was relatively low ($n = 10$ for smartphones and $n = 6$ for vacuum cleaners).

3.2.3. Best- and worst-case interpretations

The difference in interpretations when scoring on specific aspects was checked to determine the reliability of different scoring systems. This difference in best- and worst-case scenario provides an indication of reliability within each criterion.

Since the data is paired (same products used between the scoring systems) and is not normally distributed, Wilcoxon Signed-Rank Test is conducted between the best- and worst-case scenarios to determine the significance of differences between the scores.

3.3. Analysis of ease of disassembly assessment method

This section outlines the method employed for analyzing two different methods to assess ease of disassembly, proxy time method (DaRT) and step method. This was carried out using data obtained on large scale disassembly of products in professional testing labs experienced in disassembly of the products.

3.3.1. Products and its priority parts for investigation

Four washing machines, five vacuum cleaners and five smart TVs were evaluated as the result of a collaboration with consumer organisations in the framework of the European PROMPT project based on large variation in design features (Hann, 2022). Investigated priority parts are as follows:

1) **Vacuum cleaner:** suction Hose, Dust cover, Handle, Cord Reel, Motor, On/Off Switch, Wheels
2) **Televisions:** Main Board, Tcon board, Display assembly, Internal Power supply

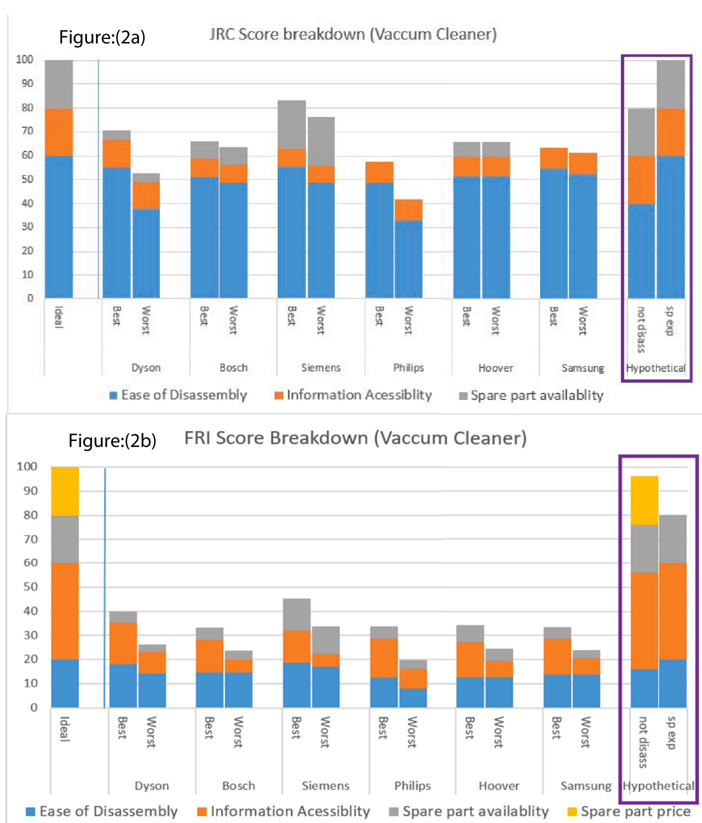

Fig. 2. Score breakdown for vacuum cleaners. 2a = JRC Score, 2b = FRI score, "sp exp " = expensive spare part "; not disass" = *a* priority part cannot be disassembled.

3) **Washing machine:** Door lock, Door seal, Electronics, Hoses, Pump, Shock Absorbers, Tub Assembly.

These priority parts are chosen based on high failure likelihood and high functional relevance.

3.3.2. *Protocol for ease of disassembly assessment method*

1. First, the products were disassembled using the disassembly protocol described in Section 3.1.
2. The number of disassembly steps and DaRT proxy times were extracted from the disassembly map.
3. The product was disassembled a second time following the third step of the disassembly protocol. However, the researcher did not think out loud during the disassembly process.
4. The entire disassembly process was recorded on video.
5. The actual times needed for disassembling (and putting aside) each priority part was extracted from the recorded video.

3.3.3. *Analysis*

Regression analysis was conducted to find the relation of actual disassembly time versus disassembly steps and DaRT time.

4. Results

4.1. *Scoring results of tested products*

The scores for the smartphones, including a breakdown of the scoring criteria, as determined using the scoring systems of JRC, FRI, and iFixit are depicted in Figs. 1a, b, and c respectively. For each product the result of best- and worst-case interpretation of scoring criteria is shown. These figures also show in the first column the ideal scenario result which is the maximum score achievable, while the last columns (highlighted in purple) show the scores for the hypothetical scenarios in which repair is considered ideal except for either a spare part that is prohibitively expensive or a priority part that cannot be disassembled. Similarly, the scores for the vacuum cleaners, including a breakdown of the scoring criteria, as determined using the scoring systems of JRC and FRI, are shown in Figs. 2a, and b, respectively.

Comparison of scores of the same smartphones obtained with the different scoring systems is presented in Fig. 3a. Comparison of scores of the same vacuum cleaners obtained with the different scoring systems is presented in Fig. 3b. The exact breakdown of the all the scores is provided in the supplementary material 9.3.

Results for the ONR system are not shown. ONR has 11 must-have criteria, for which none of the assessed vacuum cleaners passed. Essentially, the overall score of ONR for all vacuum cleaners was 0. Furthermore, scoring of vacuum cleaners by ONR turned out to be

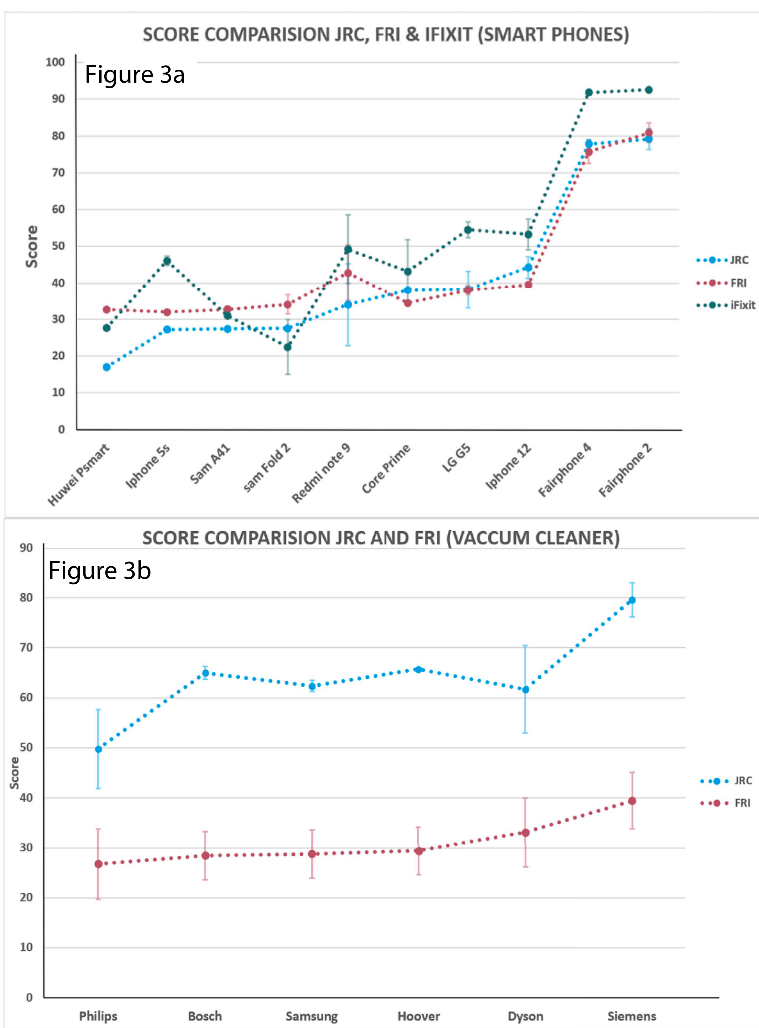

Fig. 3. Score comparison for the different scoring systems and their uncertainty range. Bars show the extent of best- and worst-case interpretations. 3a: FRI, iFixit and JRC score for smartphones. 3b: JRC and FRI scores.

highly subjective. This subjectiveness arises because ONR frequently provides 0–10 score levels but does not provide instructions on how each level should be scored. For example, one of the criteria states, "The appliances should be mostly reducible to individual components, which should also be available as spare parts," with scoring ranging from 1 to 10. However, terms like "mostly reducible" and "individual components" are open to interpretation. Additionally, there are no instructions on how it should be scored if the appliance is reducible to individual components, but spare parts are not available. This subjectivity was also observed in previous research (Bracquené et al., 2018, Dangal et al., 2022).

4.2. Performance of hypothetical scenarios

For the hypothetical scenario, for both smartphones and vacuum cleaners, in which one spare part is prohibitively expensive in a further ideally repairable product, high scores were obtained. JRC and iFixit scored up to 100 % (Figs. 1a, 1c, 2a & 2b) while FRI scored up to 80 % (Figs. 1b & 2b). Similarly, for the hypothetical scenario, in which a single priority part cannot be disassembled, FRI and JRC exceeded 80 %, and iFixit scored 68 %.

4.3. Comparison of scoring systems

For smartphones, regression analysis showed that JRC, iFixit and FRI scores were highly correlated ($R^2 = 0.98$). However, this correlation does not extend to the criteria level, where a lower score in one criterion is often offset by a higher score in another. FRI has higher scores in disassembly ease compared to JRC due to its emphasis on fewer priority parts (4 for FRI versus 8 of JRC) and less stringent disassembly steps (for example for the disassembly of the display, JRC requires less than 3 steps for the highest score, whereas the FRI requires less than 7 steps). Additionally, FRI scores high to criteria such as software reset, update information, and remote assistance availability, which are not considered in the JRC system. However, FRI includes spare part price, and this results in lower scores for most smartphones since most (8/10) score 0 in this criterion. JRC scoring system, despite having stricter disassembly criteria, tend to lean towards higher scores due to the significant weight given to the ease of disassembly (55 % for JRC versus 20 % for FRI). iFixit scores are higher than both JRC and FRI as it focuses on two easily removable key parts (battery and screen), has highest weight in ease of disassembly (58 %), and factors in the availability of third-party spare parts, which are usually easy to acquire.

For vacuum cleaners, regression analysis showed a low correlation ($R^2 = 0.59$). Wilcoxon signed-rank test confirmed that the FRI score of vacuum cleaners is significantly lower than the score of JRC ($P < 0.0001$). In the case of vacuum cleaners, the significantly lower score on FRI than JRC can be attributed to three major factors. The primary factor is the high weight assigned to JRC's disassembly criteria (60 %) in contrast to FRI's disassembly criteria (20 %). The JRC-score for disassembly only exceeds for all product the total FRI-score. Further, in the JRC system, spare part availability distinguishes between "available", 5 years, and 8 years, whereas FRI distinguishes between 9, 11, and 13 years. As a result, nearly all vacuum cleaners (5 out of 6) receive poor scores for spare part availability in the FRI system. Finally, JRC does not consider the price of spare parts as a criterion, while all vacuum cleaners score 0 in the FRI system in this respect.

4.3.1. Best- and worst-case interpretations

As presented from the Fig. 1 and Fig. 2, and based on Wilcoxon signed rank test, significant difference between worst and best interpretations are observed in both vacuum cleaners ($P < 0.001$) and smartphones ($P < 0.004$). This means that the variance from best case to worst case within the same product is higher than the difference between product scores, especially for vacuum cleaners. The main reasons for the observed differences between best- and worst-case interpretation are listed in Table 2. These differences arise from unclarity related to bundling, breakable fasteners, adhesive removal, spare part price, spare part availability, and access to repair information.

4.4. Assessment of ease of disassembly

Fig. 4 represents the number of disassembly steps and the calculated DaRT proxy time compared to the actual recorded time in professional disassembly. Based on the overall data set, step values were multiplied by 26.3 s/step (actual time mean/average steps mean) to obtain the best fit trendline with actual time.

Considering the data for all three product categories (vacuum cleaner, washing machine, smart TV), a high correlation ($R^2 = 0.98$) was observed between DaRT proxy time and actual time, and medium correlation ($R^2 = 0.86$) between number of steps and actual time. The R^2 relation and trendline for each product is presented in supplimentary material 9.5. While DaRT proxy time has a high correlation with the actual time for all product categories, the correlation between steps and actual time is quite poor for the TV and vacuum cleaner (see Fig. 3b); only its high correlation for washing machines brings up the overall correlation. Furthermore, the intercept of the trendlines for steps deviates significantly from the expected 0 intersection, which is not the case for DaRT proxy time. Finally, the best fit to all data causes the step-based times for vacuum cleaners and washing machines to be significantly overestimated. The step method only provides a reasonable indication of disassembly time for washing machines.

The deviations with the number of steps as a measure for the actual disassembly time are due to two main reasons: a) number of fasteners removed with same tool does not influence the step count, however, most of the time tool has to be changed frequently, and b) some products have many easy to remove small components, resulting in a high disassembly step count, which, nevertheless, does not take that much time.

In general, DaRT proxy time is slightly higher than the actual time. Upon video observation, the following two main reasons were identified. First, the repairer frequently keeps small tools still in hand while conducting minor disassembly tasks, while DaRT considers every change from tool to hand and vice versa. Second, removal of several components or fasteners is sometimes done at once, whereas in DaRT each removal of a component or group of fasteners is considered individually.

5. Discussion

The validity of scoring systems in hypothetical scenarios, upon comparing different scoring systems and on different disassembly assessment methods, is discussed first. Then, we discuss the reliability of the scoring systems in best- and worst-case scenarios. In general, all the assessed scoring systems have opportunities for improvement regarding their validity and reliability.

The scoring systems reflected the validity well for products that are known to be highly repairable (such as Fairphone). However, we observed validity problems across all scoring systems in scenarios, where the repair of a specific priority part was not feasible or affordable. Here, products still scored relatively high, despite poor repairability. Therefore, in these cases, none of the evaluated scoring systems accurately represents the product's actual reparability. Although these scenarios are not frequent, they do happen regularly, corroborating previous findings (Barros and Dimla, 2023, HOP, 2020).

A potential solution to address this issue could involve the implementation of a limiting factor approach (Bautsch, 2010) for criteria that determine the feasibility of repair. In this approach, if a criterion fails to meet a specified threshold, the overall score would be predominantly determined by this critical criterion, based on how far it falls below the threshold. Other recommended solutions include pass/fail thresholds and increasing assigned weight (HOP, 2020). However, the limiting factor approach ensures that other criteria are not underestimated in the

Table 2
Main reason for difference in score due to difference in interpretation and its explanation.

Design aspect	Interpretation issue in scoring	Possible scoring decisions	Advantage	Disadvantage	Affected scoring system
Bundling	Unclarity on how bundling should be addressed.	Bundled components are considered not removable	Bundling is penalised	Does not consider small elements bundled together	FRI, JRC, iFixit
		Consider the entire bundle removable	Considers small elements bundled together	Could promote bundling of large modules	
		Remove the part destructively		Promotes designs requiring destructive disassembly	
Breakable fasteners	Fasteners (e.g., snap fits) that may break during disassembly/reassembly is difficult to consider.	Consider not removable	Promotes fastener that do not break during disassembly	Products may be harshly scored since it's only a chance.	FRI, JRC, iFixit
		Consider removable	Conforms to what is currently done in general for FRI	Promotes fastener that break during disassembly	
Adhesive removal process	Multiple tools and techniques may be required, making it difficult to determine the best way to represent the adhesive removal process	Count each and every tool changed by assessor	Can address differences in difficulties of models	Is subjective	FRI, JRC, iFixit
		Determine and consider standard amount of tool change	Provides consistent and objective assessment for adhesive removal	Database and further research are required	
	Components are generally glued in thin lines, therefore unclear how to check glued surfaces	Assess entre surface that is attached	Easier to test	Less accurate on representation on the ease	iFixit
		Assess only surfaces where glue is present	More difficult to test	More accurate on representation on the ease	
Spare part availability	Unclear how "out of stock" spare parts should be handled	Consider not available	Easier to test	Do not account for temporary out of stock parts	FRI, JRC, iFixit
		Contact manufacturer when it will be in stock	More representative of real case	Difficult to get answers from Manufacturer	
Repair information	Unclear the extent of information required for a "pass" criterion.	Thorough step by step information for diagnosis and repair on each failure mode is required	More accurate representation of information availability	Maybe too strict and currently none of the products will pass. Do not consider partial information	JRC, FRI, iFixit
		Partial guidance on general failure mode may suffice	Points are provided	Less strict and can be easily bypassed without giving clear guidance to user	

absence of detrimental features and avoids harsh penalties for products that slightly exceed the threshold value. For instance, if the threshold value is set at 30 %, a product with a score of 29 % would face severe penalties, while one with a score of 31 % would not at all.

This limiting factor approach also allows for balanced distribution of criteria. Currently, The Joint Research Centre (JRC) prioritizes disassembly efficiency, indicating that ease of disassembly enables quicker, safer, and less costly repairs, thereby facilitating better accessibility for both professional and amateur repairers (Cordella et al., 2019). However, here other critical factors such as spare part availability are given low weight. In contrast, the French Reparability Index (FRI) assigns a lower weight to disassembly and focuses more on factors such as spare part availability and price. These differing weights can lead to scenarios where a high score in one criterion compensates for a poor score in another, undermining the overall goal and validity of assessing reparability. The limiting factor approach prevents this, while ensuring important aspects such as ease of disassembly are not overshadowed.

Currently, only FRI takes spare part price into account (with a weight of 20 %), resulting in situations where expensive repairs make repair unfeasible to the customers despite receiving a high score. Spare part price is usually not accounted for because it is hard to assess precisely, as prices fluctuate over time, but such an approach fails to capture critical data, as a result, this is also identified as one of the main issues related to reliability. Possible solutions include using price estimates with large error margins, asking manufacturers to provide information on how spare part prices would change over time, or conducting regular price checks. However, each method has its limitations, and further research may be necessary to effectively tackle this issue.

Software aspects are addressed by two scoring systems (JRC and FRI); for smartphones, software has a weight of 15 % and 5 %, respectively. For devices centred around software (e.g., smartphones and laptops), long term support of security and software updates are crucial for extended lifetime of the device and could be a deciding factor for users to repair when the product eventually fails (Van den Berge et al., 2021; Jacobs and Hörisch, 2022). Additionally, software part-pairing, where a part becomes difficult or almost impossible to replace by third party repairers or users, is becoming increasingly prevalent (HOP, 2020). However, none of the scoring systems take this into account. Thus, while this study did not empirically test software aspects, software update time and part pairing should be investigated to establish if software should be represented with a higher weight or considered as a limiting factor.

To determine priority parts, according to recent research (Bracquené et al., 2021) considering either 75 % of failures or at least 5 priority parts provides the optimum balance between validity and ease of testing. In situations where the failure rates of priority parts are significantly different (for example, 60 % battery, 20 % motor for cordless vacuum cleaners), the validity of the scoring system could be improved if individual scores related to priority parts (e.g., ease of disassembly of a priority part, spare part availability) are weighted according to the average failure rate per product category (Bracquené et al., 2021). Given this, the JRC scoring system (with weights based on failure rates) appears to be most valid, followed by FRI. The iFixit scoring system considers only the battery and screen as priority parts, and the combined failure rates for these two are approximately 60 % (Cordella et al., 2020). Thus, the iFixit scoring system could be enhanced by incorporating one or two additional priority parts in the assessment.

Fig. 3 shows that the proxy time method provides a better representation of the ease of disassembly than the number of steps. Specifically, DaRT proxy times are valid for a wide range of product categories, as demonstrated by their validity for washing machines, TVs, and vacuum cleaners, while number of steps were only valid for washing machines. This largely extends the validated applicability of proxy times for assessing ease of disassembly. Since ease of disassembly is one of the

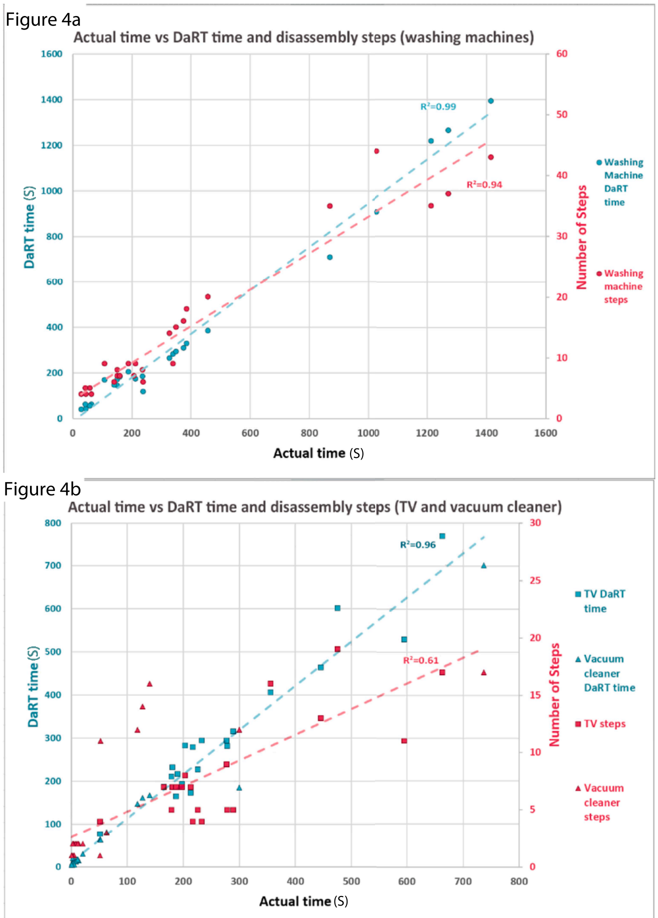

Fig. 4. Calculated DaRT times and disassembly steps compared to actual disassembly times. For calculating the correlation coefficients, a step is assumed to take 26.3 s on average. 4a: washing machine data with. 4b: vacuum cleaner and smart TV data.

critical criteria for determination of reparability, a proxy time method is recommended for a higher validity scoring system.

Reliability findings, as presented in Table 2, show that while there is reasonable correlation between scores of the same product across scoring systems, each product's possible variation of score from best to worst was often larger than the differences between products; this is in line with previous research (Barros and Dimla, 2023, HOP, 2020). Thus, scoring systems need clearer and more detailed protocols to minimize subjectivity. One major reliability issue relates to bundling. Here, clearer protocols are required so that designs containing large modules or requiring destructive disassembly are adequately penalized. For example, bundled parts might be considered as non-removable. However, it is important to note that the use of bundling can also be beneficial for the ease of disassembly, as well as the product's reliability and cost (Dangal et al., 2022). Striking the right balance when applying bundling is crucial, and this balance should be reflected in the scoring accordingly. Other reliability issues relate to "repair information" and "diagnosis information" criteria, which are not accurately described by JRC and FRI. The level of detail in diagnosis and repair information should be specified by indicating exact failure modes and parts it addresses steps to diagnose and repair it.

5.1. Future opportunities

Validity can be further examined by surveying professional repairers about the product's reparability and evaluating how well their opinions align with the score. Furthermore, comparison of disassembly step numbers and DaRT times versus actual times for glued components should be assessed to provide a complete picture on the validity of each disassembly assessment method.

The reliability of scoring systems can be further improved by engaging a larger number of independent assessors to evaluate the scoring system using the same products. Evaluating more product categories would help to determine the scoring systems' generalizability across different categories.

Reparability index, with indexes also integrating durability

6. Conclusion

This study evaluated the validity and reliability of four prominent reparability scoring systems: the French Reparability index (FRI), Joint Research Centre scoring system (RSS or JRC scoring system), iFixit, and ONR19202. The research aimed to enhance the effectiveness of these systems in promoting more durable and repairable products in alignment with the European Commission's circular economy action plan and the right to repair movement.

While the scoring system provided a valid reflection of products known to be repairable, the analysis also identified several areas that require improvement. Scenarios where repair is deemed infeasible or too expensive demonstrated that the current scoring systems do not accurately represent the actual reparability of products. To address this issue, the study proposed the implementation of a limiting factor approach for criteria that determine the feasibility of repair. This approach ensures that critical criteria play a significant role in determining the score when they fail to meet specified thresholds.

Reliability issues in the scoring systems were also identified. Different interpretations of scoring criteria can change overall product scores significantly, making them score better or worse than other products. The paper recommends a need for a clearer and more detailed protocol to address issues related to bundling, spare part availability and price, and repair and diagnosis information criteria, ensuring that these design aspects are free from subjective interpretations.

Additionally, the study evaluated the accuracy of disassembly assessment methods, specifically the DaRT method based on proxy times (which estimates disassembly times by assigning proxy times to actions) and the step method (which counts the number of disassembly steps), in comparison to actual disassembly time. Results indicated that the DaRT proxy time method gives a better representation of actual disassembly ease than the step method. It accurately assessed washing machines, vacuum cleaners, and TVs, while the steps method was only accurate for washing machines, and even for these it was less accurate than the DaRT method.

Our findings can help to improve scoring systems, which in turn should improve product design for sustainability. The suggestions presented could serve as resources for designers, repair professionals, consumer organizations, and policymakers, aiding them in promoting more repairable products. For repair scoring systems to be widely adopted in policy or industry, ensuring their validity and reliability is essential to establish trust and credibility.

CRediT authorship contribution statement

Sagar Dangal: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Sonia Sandez:** Formal analysis, Data curation. **Julieta Bolaños Arriola:** Formal analysis, Data curation. **Jeremy Faludi:** Writing – review & editing, Supervision, Methodology. **Ruud Balkenende:** Writing – review & editing, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was funded by the European Commission under the Horizon 2020 Premature Obsolescence Multi stakeholder Product Testing Program (PROMPT) (Grant Agreement number 820331).

The research visit of Sonia Sandez was funded by the Consellería de Innovación, Universidades, Ciencia y Sociedad Digital, Program for the promotion of scientific research, technological development and innovation in the Valencian Community. (Spain) (BEFPI-2021) and Generalitat Valenciana (Spain) (ACIF/2020/334)

We would like to sincerely thank Dr. Linda Ritzen for her valuable assistance with the analysis and verification of the DaRT method.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.resconrec.2025.108211.

Data availability

Data will be made available on request.

References

Bannigan, K., Watson, R., 2009. Reliability and validity in a nutshell. J. Clin. Nurs. 18 (23), 3237–3243. https://doi.org/10.1111/j.1365-2702.2009.02939.x.

"Repairability Index," Ministère De La Transition Énergétique. 2024 [Online]. Available: https://www.ecologie.gouv.fr/indice-reparabilite.

Barros, M. and Dimla, E., "Smartphone repairability indexes in practice Linking repair scores to industrial design features," pp. 1–14, 2023, doi: 10.1111/jiec.13398.

Boix, N., Gabriel, C., Gaha, R., Favi, C., 2023. ScienceDirect Analysis of disassembly parameters in repairability scores : limitations for engineering design and suggestions for improvement. Procedia CIRP. 116, 738–743. https://doi.org/10.1016/j.procir.2023.02.124.

Bracquené, E., et al., 2018. ASMER BENELUX Reparability Criteria for Energy Related Products Study in The Benelux Context to Evaluate the Options to Extend the Product Life Time. BeNeLux [Online]. Available. http://www.benelux.int/files/7915/2896/0920/FINAL_Report_Benelux.pdf.

Bracquené, E., Peeters, J., Alfieri, F., Sanfelix Forner, J.V., Duflou, J., Dewulf, W., Cordella, M., 2021. Analysis of evaluation systems for product repairability: a case study for washing machines. J. Clean. Prod. 281, 125122. ISSN 0959-6526, JRC122689.

C, S., et al., 2022. Product Reparability Scoring System: Specific application to Smartphones and Slate Tablets. Publications Office of the European Union, Luxembourg. https://doi.org/10.2760/340944. Luxembourgonline.

Cordella, M., Alfieri, F., Sanfelix, J., 2019. Analysis and development of a scoring system for repair and upgrade of products. EUR 29711 EN, Publications Office of the European Union, Luxembourg. https://doi.org/10.2760/725068. ISBN 978-92-76-01602-1, JRC114337.

Cordella, M., Alfieri, F., Sanfelix, J., 2020. Guidance for the Assessment of Material Efficiency: Application to Smartphones. https://doi.org/10.2760/037522.

Dangal, S., Faludi, J., Balkenende, R., 2022. Design aspects in repairability scoring systems: comparing their objectivity and completeness. Sustainability 14 (14), 8634. https://doi.org/10.3390/su14148634.

De Fazio, F., Bakker, C., Flipsen, B., Balkenende, R., 2021. The disassembly map: a new method to enhance design for product repairability. J. Clean. Prod. 320 (April 2020), 128552. https://doi.org/10.1016/j.jclepro.2021.128552.

Indice De Réparabilité, 2022. Accessed: Apr. 19[Online]. Available. https://www.ecologie.gouv.fr/indice-reparabilite.

EN45554, 2020. General methods for the assessment of the abilitLast sync: Last synced 0 min ago Sagar Dangal's user image Sagar Dangal Yellow 834 Info Annotations Notebook REPORTy to repair, reuse and upgrade energy-related products.

UNEP Circularity Platform, 2024. The French Approach to Circular Economy and Coherent Product Policies. [Online]. Available https://buildingcircularity.org/the-french-approach-to-circular-economy-and-coherent-product-policies/#:~:text=France%20is%20the%20first%20country,2023%20for%20other%20product%20categories.

European Commission, 2020. New Circular Economy Action Plan For a Cleaner and More Competitive Europe, pp. 1–23 [Online]. Available. https://www.europarl.europa.eu/news/en/headlines/economy/20151201STO05603/circular-economy-definition-importance-and-benefits?&at_campaign=20234-Economy&at_medium=Google_Ads&at_platform=Search&at_creation=RSA&at_goal=TR_G&at_audience=circular-economy-acti.

Flipsen, B., Bakker, C., Van Bohemen, G., 2016. FLIPSEN developing a reparability indicator for electronic products. In: 2016 Electronics Goes Green 2016+, EGG 2016, pp. 1–9. https://doi.org/10.1109/EGG.2016.7829855.

Flipsen, B., Huisken, M., Opsomer, T., Depypere, M., 2019. IFIXIT smartphone reparability scoring: assessing the self-repair potential of mobile ICT devices. In: PLATE Conference 2019, pp. 18–20.

Bautsch, M., "On Limiting Effects in Comparative Testing," 2010. 10.13140/RG.2.2.35313.25446.

Hervier Marie, A., De, et al., "Benchmark International," 2018.

Hann, D., 2022. Premature obsolescence multi-stakeholder product testing program: generalization of approach and summary of results. https://cordis.europa.eu/project/id/820331/results.

HOP, 2020. The French Repairability Index Executive Summary.

iFixit, 2021. Smartphone Repairability Scores. Accessed: Aug. 02[Online]. Available. https://www.ifixit.com/smartphone-repairability.

Jacobs, K. and Hörisch, J., "The importance of product lifetime labelling for purchase decisions : strategic implications for corporate sustainability based on a conjoint analysis in Germany," no. October 2021, pp. 1275–1291, 2022, doi: 10.1002/bse.2954.

Marikyan, D., Papagiannidis, S., 2024. Exercising the 'Right to Repair': a customer's perspective. Journal of Business Ethics 193 (1), 35–61. https://doi.org/10.1007/s10551-023-05569-9. Aug.

Ministère de la Transition Écologique, 2024. Indice De Réparabilité - Appareils [Online]. Available. https://www.indicereparabilite.fr/appareils/.

ONR 192102, 2014. ONR 192102 Label of Excellence for Durable, Repair Friendly, Designed Electrical and Electronic Appliances [Online]. Available. http://step-initiative.org/index.php/WorldMap.html.

Paper, H.I.L.B.C., "Validity and reliability in qualitative research," vol. 16, no. 2, pp. 35–38, 1993.

Schischke, K., Berwald, A., Dimitrova, G., Rückschloss, J., Nissen, N.F., Schneider-Ramelow, M., 2022. Durability, reparability and recyclability: applying material efficiency standards en 4555x to mobile phones and tablet computers. Procedia CIRP. 105, 619–624. https://doi.org/10.1016/j.procir.2022.02.103.

Van den Berge, R., Magnier, L., Mugge, R., 2021. Too good to go? Consumers' replacement behaviour and potential strategies for stimulating product retention. Curr. Opin. Psychol. 39 (July), 66–71. https://doi.org/10.1016/j.copsyc.2020.07.014.

Vanegas, P., et al., 2018. Ease of disassembly of products to support circular economy strategies. Resour. Conserv. Recycl. 135 (January 2017), 323–334. https://doi.org/10.1016/j.resconrec.2017.06.022.