



# Wie wählt man einen guten, aber praktikablen Fragebogen?

## Empfehlungen auf Basis der Kursevaluation der BYTE Challenge

Clarissa Sabrina Arlinghaus<sup>1,2</sup>

Angenommen: 2. Januar 2024  
© The Author(s) 2024

### Zusammenfassung

Die BYTE Challenge ist eine kostenlose Online-Lernplattform für Schüler:innen ab Klasse 5. Zur Evaluation des Kursangebots wird ein starker, quantitativer Ansatz verfolgt. Bisher waren die Auswertungsmöglichkeiten des Teilnehmenden-Feedbacks aufgrund ordinalskaliertter Antwortoptionen jedoch auf deskriptive Statistiken beschränkt. Um künftig aus einem breiteren Spektrum an Auswertungsmethoden schöpfen zu können, wurde der Fragebogen mit intervallskalierten Antwortoptionen überarbeitet. Die Überarbeitung des Fragebogens vereinbart theoretische Ansprüche und gewünschte Praktikabilität. In diesem Beitrag werden die in der BYTE Challenge intern gesammelten Überlegungen und Erfahrungswerte mit der Öffentlichkeit geteilt. Darüber hinaus werden anwendungsorientierte Empfehlungen für die auf Fragebögen basierende Evaluation anderer Praxisprojekte gegeben.

Die BYTE Challenge ist eine kostenlose Online-Lernplattform für Schüler:innen ab Klasse 5 mit über 100 MINT-Kursen in verschiedenen Schwierigkeitsstufen und drei Sprachen (deutsch, englisch, ukrainisch). Abgedeckt werden Themen wie z. B. Social Media, Datenschutz, Digitale Cookies, Algorithmen & Datenstrukturen, Programmierung mit Scratch, künstliche Intelligenz [1].

Die BYTE Challenge ist ein Projekt der Gesellschaft für Informatik e. V. und wurde im Mai 2020 von Carolin Neumann und Stefan Hildebrand gegründet [2]. Bereits 2021 konnte die BYTE Challenge über 1000 Teilnahmen verzeichnen. Von über 800 Teilnahmen aus über 80 Schulen lagen genug Daten vor, um eine Evaluation der BYTE Challenge durchzuführen. Durch unsere Auswertung konnten wir anhand der erreichten Punktzahlen belegen, dass unser Lernangebot für Schüler:innen von Gymnasien und Oberschulen gleichermaßen geeignet ist und unterschiedliche Vorkenntnisse in Scratch ausbalanciert werden können [3].

Zwar verfolgten wir hier bereits einen starken, quantitativen Ansatz, aber waren in der Auswertung des Feedbacks zu einzelnen Kursen auf deskriptive Statistiken beschränkt [3]. Der Grund hierfür war die Verwendung von ordinalskalierten Antwortoptionen im Feedback-Fragebogen (siehe Abb. 1). Ordinalskalierte Daten lassen sich der Größe oder Wertigkeit nach ordnen, aber verfügen nicht über gleichmäßige Abstände wie intervallskalierte Daten. Dies hat zur Folge, dass einige Testverfahren (z. B. t-Tests oder Varianzanalysen) nicht zur Datenauswertung eingesetzt werden können, da hierfür die abhängige Variable ein Intervallskalenniveau aufweisen muss [4].

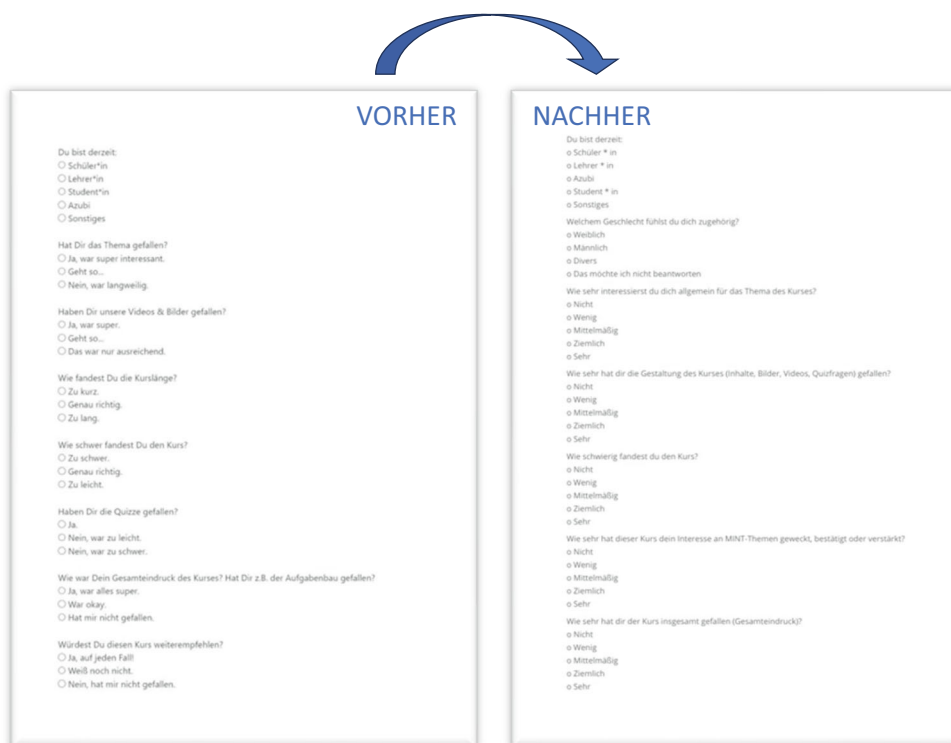
Um künftig aus einem breiteren Spektrum an Auswertungsmethoden schöpfen zu können, entschieden wir uns für die Überarbeitung des Feedback-Fragebogens. Im neuen Feedback-Fragebogen sollte überall ein Intervallskalenniveau vorliegen. Dies hätte man z. B. dadurch erzielen können, indem man alle Fragen auf einer Skala von 0 bis 100% beantworten ließe. Für manche Zwecke mag das angemessen sein, doch in unserem Fall hielten wir es für nicht geeignet. Vermutlich fällt es Schüler:innen schwer, ihr Interesse an einem Thema oder ihre Kursbewertung auf einer Skala von 0 bis 100% einzuschätzen. Statt einer erhofften stärkeren Differenzierbarkeit würden wir aufgrund von kognitiver Überforderung wahrscheinlich nur wilde Schätzungen mit fraglicher Aussagekraft erhalten. Aus diesem Grund bevorzugen wir ein Antwortformat mit weniger Ausprägungen.

✉ Clarissa Sabrina Arlinghaus  
clarissa.arlinghaus@gi.de

<sup>1</sup> Universität Bielefeld, Bielefeld, Deutschland

<sup>2</sup> BYTE Challenge, Gesellschaft für Informatik, Berlin, Deutschland

Abb. 1 Überarbeitung des BYTE-Evaluationsfragebogens



Empfohlen wird eine fünf- bis siebenstufige Antwortskala zur Maximierung der Reliabilität der Messwerte [5]. Bei fünf oder mehr Abstufungen kann die Skala annähernd als intervallskaliert betrachtet werden, auch wenn sie strenggenommen ordinalskaliert ist [6]. Dies sollte bei der Auswahl von Fragebogen-Items berücksichtigt werden. Idealerweise sollten bereits erprobte, validierte, intervallskalierte Fragebögen genutzt werden. Sofern man ein Produkt bewertet, empfehlen sich hier Fragebögen wie beispielsweise der meCUE [7], AttrakDiff [8] oder UEQ-S [9]. Was aber, wenn man kein Produkt bewertet, sondern eine Dienstleistung? Dann kommt z. B. der Net Promoter Score infrage, der die Weiterempfehlungswahrscheinlichkeit misst [10]. Auch zur Evaluation von Lernangeboten existieren Fragebögen wie z. B. der FLIPPY [11] oder das TEI [12].

Obwohl alle genannten Fragebögen eine gute Qualität aufweisen, fanden wir sie nicht passend für die Evaluation der BYTE Challenge, da wir unseren Schwerpunkt nicht auf die User Experience [7–9] legen und eine Zielgruppe

haben, die zu jung ist, um z. B. nach der Anwendbarkeit im beruflichen Alltag [12] gefragt zu werden. Zudem schätzen wir den FLIPPY [11] oder das TEI [12] als zu lang ein für unsere Kurse, die Schüler:innen kostenlos und unentgeltlich in ihrer Freizeit absolvieren. Hier vermuteten wir eine zu hohe Abbruchrate, die zu sehr lückenhaften und damit schwer verwertbaren Datensätzen führen könnte. Außerdem wollen wir gerne zukünftige Evaluationsergebnisse mit der Evaluation vom Jahr 2021 [3] vergleichen können. Daher war es uns wichtig, die gleichen Bereiche (Interesse, Schwierigkeit, Inhalt, Gesamteindruck) [3] anzusprechen.

Aus diesem Grund entschieden wir uns dafür, eigene Fragebogen-Items zu entwickeln. Schnell befanden wir uns in einem Zwiespalt zwischen theoretischen Ansprüchen und gewünschter Praktikabilität. Uns war es wichtig, am Intervallskalenniveau und der Kürze unseres bisherigen Fragebogens festzuhalten. Daher entschieden wir uns als Praxislösung dafür, auf die Antwortskalen von Rohrmann [13] zurückzugreifen (siehe Tab. 1 und Abb. 1). Hierbei handelt es sich um einfache, leicht verständliche Antwortoptionen in deutscher Sprache. Der große Vorteil besteht darin, dass die Antwortoptionen trotz ihrer verbalen statt numerischen Darstellung und auch bei der Verwendung als einzelne Items annähernd ein Intervallskalenniveau aufweisen [13]. Dank des Intervallskalenniveaus können wir also demnächst auch das Feedback mit inferenzstatistischen Methoden (z. B. t-Tests oder Varianzanalysen) auswerten [4].

Ferner sollten keine unipolaren (z. B. 1, 2, 3, 4, 5) und bipolaren Skalen (z. B. -2, -1, 0, +1, +2) miteinander ver-

Tab. 1 Antwortskalen von Bernd Rohrmann

<i>Häufigkeit</i>				
Nie	Selten	Gelegentlich	Oft	Immer
<i>Intensität</i>				
Nicht	Wenig	Mittelmäßig	Ziemlich	Sehr
<i>Bewertung von Aussagen</i>				
Stimmt nicht	Stimmt wenig	Stimmt mittelmäßig	Stimmt ziemlich	Stimmt sehr

mischt werden, da dies die Interpretation weniger eindeutig macht [6]. Am einfachsten ist es für Befragte, wenn allen Fragen die gleiche Antwortskala zugrunde liegt [13]. Daher formulierten wir unsere Feedback-Fragen so, dass wir immer Rohrmanns Skala zur Messung von Intensität [13] heranziehen konnten.

Die Frage nach der aktuellen Lebenssituation behielten wir dennoch bei, um unsere Stichprobe beschreiben und zu Evaluationszwecken filtern zu können. Zudem haben wir eine Frage nach der Geschlechtszugehörigkeit ergänzt, um in Zukunft mögliche Geschlechtsunterschiede näher beleuchten zu können.

Wir empfinden das als eine gelungene und trotzdem praxisnahe Verbesserung unseres Feedback-Fragebogens und möchten unsere Erfahrungen und Überlegungen mit anderen teilen. Aus diesem Grund präsentieren wir in dieser Ausgabe unseren neuen Feedback-Fragebogen (siehe Abb. 1).

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Open Access** Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

## Literatur

1. Challenge BYTE (n.d.). Website der BYTE Challenge. <https://byte-challenge.de>
2. Hildebrand S, Neumann C (2021). Die BYTE Challenge – One bit at a time. In INFORMATIK 2021. Gesellschaft für Informatik, S 1641–1644 <https://doi.org/10.18420/informatik2021-138>
3. Arlinghaus CS, Hildebrand S, Neumann C (2022) BYTE Challenge—from competition to STEM platform. In Proceedings of the 17th Workshop in Primary and Secondary Computing Education. WiPSCe, Bd. 22. ACM, Artikel 9, S 1–6. <https://doi.org/10.1145/3556787.3556863>
4. Bortz, J. & Schuster, C. (2010). Statistik für Human- und Sozialwissenschaftler. (7., vollständig überarbeitete und erweiterte Auflage mit 70 Abbildungen und 163 Tabellen). Springer. <https://doi.org/10.1007/978-3-642-12770-0>
5. Bühner M (2010) Einführung in die Test- und Fragebogenkonstruktion. Pearson Studium
6. Moosburger H, Kelava A (2020) Testtheorie und Fragebogenkonstruktion, 3. Aufl. Springer <https://doi.org/10.1007/978-3-662-61532-4>
7. Minge, M & Riedel, L. (2013). meCUE – Ein modularer Fragebogen zur Erfassung des Nutzungserlebens. In S. Boll, S. Maaß & R. Malaka (Hrsg.), Mensch und Computer 2013: Interaktive Vielfalt. Oldenbourg Verlag, S 89–98. [https://doi.org/10.1524/9783486781229\\_89](https://doi.org/10.1524/9783486781229_89)
8. Hassenzahl M, Burmester M, Koller F (2003) AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In: Szwillus G, Ziegler J (Hrsg) Vieweg + Teubner Verlag. Mensch und Computer 2003. Berichte des German Chapters of the, Bd. 57. ACM, S 187–196 [https://doi.org/10.1007/978-3-322-80058-9\\_19](https://doi.org/10.1007/978-3-322-80058-9_19)
9. Schrepp, M., Hinderks, A. & Thomaschewski, J. (2017). Konstruktion einer Kurzversion des User Experience Questionnaire. In M. Burghardt, R. Wimmer, C. Wolff, C. Womser-Hacker (Hrsg.), Mensch und Computer 2017 – Tagungsband. Gesellschaft für Informatik, S 355–360. <https://doi.org/10.18420/muc2017-mci-0006>
10. Van Riet J, Kirsch M (2010) Konzeption und Nutzung des Net Promoter® Score. In: Greve G, Benning-Rohnke EE (Hrsg) Kundenorientierte Unternehmensführung. Gabler, S 36–83 [https://doi.org/10.1007/978-3-8349-8851-5\\_2](https://doi.org/10.1007/978-3-8349-8851-5_2)
11. Weber M, Gebhardt M (2022) Fragebogen zur Evaluation von Flipped Classroom Lehrveranstaltungen (FLIPPY) <https://doi.org/10.5283/epub.52251>
12. Ritzmann, Hagemann, Kluge (2020) TEI. Training Evaluations Inventar [Verfahrensdokumentation, Fragebogen Deutsch und Englisch. In: Psychologie (ZPID) (Hrsg) Leibniz Institut für ZPID, Open Test Archive <https://doi.org/10.23668/psycharchives.3454>
13. Rohrman B (1978) Empirische Studien zur Entwicklung von Antwortskalen für die sozialwissenschaftliche Forschung. Z Sozialpsychologie 9:222–245

**Hinweis des Verlags** Der Verlag bleibt in Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutsadressen neutral.