

## **Bildung, Schule, Lehren und Lernen während der Schulschließungen aufgrund der Corona-Pandemie – Ein Überblick zu bestehenden Online-Umfragen**

*Christoph Helm, Christoph Weber, Georg Krammer, Daniela Wetzelhütter*

Die Corona-Pandemie hat im März 2020 zu Schulschließungen in den deutschsprachigen Ländern (und darüber hinaus) geführt. Je nach Bundesland/Kanton wurde teils unterschiedlich vorgegangen, bspw. gab es unterschiedliche Formen der Schülerbetreuung, des Fernunterrichts, der Anwesenheit von Lehrkräften, sowie der Ferienregelungen. Diese für alle neue Situation führte rasch zu neuen Herausforderungen, vielen offenen Fragen und je nach Akteursgruppe zu unterschiedlichen Informationsbedürfnissen. Mit dem Ziel, diese Informationsbedürfnisse zu befriedigen, starteten Wissenschaftlerinnen und Wissenschaftler verschiedener Universitäten, pädagogischer Hochschulen und anderer Forschungsreinrichtungen in Deutschland, Österreich und der Schweiz Online-Umfragen. Darüber hinaus wurden Online-Umfragen auch von anderen Einrichtungen (Meinungsforschungsinstituten, Stiftungen, Kammern, Vereinen, Schulen, etc.) durchgeführt.

Da bereits eine hohe Zahl an Online-Umfragen zu unterschiedlichsten Aspekten von Bildung (Schulsystem, Schule, Unterricht, Lernen, etc.) während der Corona-Pandemie durchgeführt wurden und bereits eine Vielzahl erster (deskriptiver) Auswertungen, Forschungsberichte und Publikationen vorgelegt wurde, fällt es zunehmend schwieriger den Überblick über den Wissensstand rund um Lehren und Lernen (auf den unterschiedlichen Ebenen des Bildungssystems) während der Corona-Pandemie zu behalten. Aus diesem Grund bietet die ÖFEB auf dieser Seite eine Liste aktueller Online-Umfragen zum Thema.

Die ÖFEB möchte darauf hinweisen, dass die gelisteten Online-Umfragen keinem wissenschaftlichen Review von Seiten der ÖFEB unterzogen wurden. Die Beurteilung der wissenschaftlichen Qualität dieser Online-Umfragen muss die/der Leser/in selbst vornehmen. Gleichsam möchte die ÖFEB die/den Leser/in hier unterstützen, indem im Folgenden einige relevante Aspekte skizziert werden, die für die kritische Reflexion der wissenschaftlichen Aussagekraft derartiger Online-Umfragen zentral erscheinen (für eine umfangreichere Diskussion siehe Huber & Helm, 2020).

Vorweg: Die Ankündigung der Schulschließungen erfolgte vielerorts sehr kurzfristig. In Österreich beispielsweise wurde vom BMBWF am 11. März 2020 bekanntgegeben, dass ab 18. März 2020 bis zum Beginn der Osterferien 2020 kein regulärer Unterricht für Kinder aus Volksschulen, Neuen Mittelschulen, AHS-Unterstufen oder einer Sonderschule stattfindet. Wissenschaftlerinnen und Wissenschaftler hatten daher im besten Fall einige wenige Tage/Wochen Zeit, um Umfragen, die eine *Momentaufnahme der aktuellen Situation* liefern, zu planen und durchzuführen. Je nach Ressourcen, Erfahrung und Knowhow einer (wissenschaftlichen) Institution konnte in dieser knappen Zeit wissenschaftlichen Qualitätskriterien mehr oder weniger entsprochen werden. Im Folgenden werden einige dieser Qualitätskriterien kritisch diskutiert, wobei es sich hier nur um eine selektive Auswahl handelt. Wir weisen darauf hin, dass es sich dabei um Aspekte handelt, die nicht nur Online-Umfragen mit Corona-Bezug betreffen, sondern die gesamte Bildungsforschung bzw. die gesamten Sozialwissenschaften. Weiter sollte bedacht werden, dass in den hier gesammelten Umfragen die Daten mit sehr unterschiedlichen Zielen erhoben wurden. Manche Umfragen verfolgten stärker wissenschaftsorientierte Ziele (z.B. Informationsbasis für eine evidenzbasierte Bildungspolitik), während andere auf evaluative oder situationsbeschreibende Informationen abzielten (z.B. Befragungen für Entwicklungszwecke von Einzelschulen oder Elternverbänden). Während für erste Ziele strengere wissenschaftliche Gütekriterien anzulegen sind, sind für letztere Ziele manche Gütekriterien unter Umständen weniger bedeutsam (z.B. die theoretische Fundierung, Stichprobenumfang).

Interessierte Leserinnen und Leser, die auf einer breiteren Basis die Qualität von Umfragen bzw. Studien beurteilen wollen, sind auf das Konzept des *Total Survey Errors* verwiesen, das auf alle möglichen Fehlerquellen (Stichproben-, Spezifikations-, Abdeckungs-, Nonresponse-, Mess- und Verarbeitungsfehler) von Befragungen/Surveys fokussiert (u.a. Groves & Weisberg, 2010; Biemer & Lyberg, 2003: 37ff).

- 1. Externe Validität bzw. Verallgemeinerbarkeit der Befunde:** Die externe Validität einer Studie und somit die Verallgemeinerbarkeit von Studienergebnissen hängt wesentlich von der Repräsentativität der untersuchten Stichprobe ab. Globale Repräsentativität (d.h. in Bezug auf alle Merkmale einer Stichprobe) kann in der Regel nur in Zufallsstichproben erwartet werden, für deren Ziehung ein Stichprobenrahmen – d.h. eine Liste mit Kontaktdaten aller Einheiten (Personen, Schulen, ...) der Grundgesamtheit – nötig ist. Merkmalspezifische Repräsentativität (z.B. in Bezug auf Geschlecht oder Bildungshintergrund) wird u.a. auch durch Quoten bei der Stichprobenziehung oder durch Gewichtung (Poststratifikation, d.h. die Anpassung der realisierten Stichprobe an bekannte Merkmalsverteilungen der Grundgesamtheit) zu erreichen versucht.

Bisher sind ausschließlich Online-Umfragen bekannt, deren Ergebnisse in der Regel nicht als repräsentativ zu sehen sind. Diesbezüglich sind u.a. folgende Gründe anzuführen: a) Personen ohne (ausreichenden) Zugang zu digitalen Medien und Internet sind per se von der Befragung ausgeschlossen. b) Aufgrund von fehlenden Stichprobenrahmen werden Links zu Online-Befragungen meist über Websites oder per Mail („Schneeballverfahren“, d.h. Versenden von Mails mit der Bitte um Weiterleitung) verbreitet, wodurch sehr wahrscheinlich nicht die gesamte Grundgesamtheit erreicht wird. c) Die Teilnahmeraten an Befragungen variieren häufig in Abhängigkeit von soziodemografischen Merkmalen, wie dem Bildungshintergrund oder auch einem Migrationshintergrund bzw. der (zu Hause) gesprochenen Sprache. Durch diese Selbstselektion ist davon auszugehen, dass insbesondere jene Kinder und Familien, die vermutlich am stärksten von den Schulschließungen betroffen sind, in den bisherigen Befragungen unterrepräsentiert sind.

*Inwiefern berichtete Befunde aus den Online-Umfragen verallgemeinerbar sind, ist mit Blick auf die jeweilige Stichprobensammensetzung kritisch zu prüfen.* In der Beurteilung der hier fokussierten Umfragen sollten folgende Punkte berücksichtigt werden. I.) Wie wurde die Stichprobe realisiert? II.) Wird die realisierte Stichprobe in Hinblick auf merkmalspezifische Repräsentativität geprüft? III.) Erfolgt im Falle einer Abweichung (z.B., wenn Kinder aus bildungsfernen Familien unterrepräsentiert sind) eine Gewichtung der Daten?

- 2. Interne Validität und Kausalität:** Die interne Validität bezieht auf die Frage, ob aus Studienergebnissen Ursache-Wirkungs-Aussagen, also Aussagen über Kausalitäten abgeleitet werden können. Die interne Validität sinkt mit der steigenden Zahl an alternativen Erklärungen für ein empirisches Ergebnis. Eine wichtige – wenn auch nicht ausreichende – Voraussetzung für kausale Aussagen ist, dass sich die Ursache vor der Wirkung ereignete (und auch gemessen wird). Viele der bisherigen Online-Umfragen – insbesondere jene, die am Anfang der Corona-Pandemie – berichtet wurden, sind Querschnittsumfragen, deren Daten und Befunde keine kausalen Interpretationen zulassen (wohl aber eine Analyse der Zusammenhänge interessierender Variablen). Aktuell sind jedoch einige wenige Umfragen bekannt, die eine Messwiederholung (z.B. im Rahmen der Schulöffnungen) durchführen (bspw. die Umfragen der Universität Wien, der Universität Jena und der Pädagogischen Hochschule Zug). Kausalitäts- und Wirkungsaussagen müssen (mit wenigen Ausnahmen)

zumindest auf Paneldaten (d.h. Messwiederholungen an den gleichen Personen) basieren. Darüber hinaus sind auch immer alternative Erklärungen für Ergebnisse (z.B. durch nicht erhobene bzw. in den Analysen nicht berücksichtigte Variablen) in den Blick zu nehmen. *Kausalitäts- und Wirkungsaussagen in Berichten von Online-Befragungen sind daher mit Blick auf das Design der Umfrage kritisch zu prüfen.* Folgende Punkte sollten in der Beurteilung der Umfragen berücksichtigt werden: I.) Liegen Paneldaten vor? II.) Wurden naheliegende Drittvariablen (z.B. Variablen, die ebenso als Ursache in Frage kommen könnten) und somit alternative Erklärungen in den Blick genommen?

- 3. Reliabilität und Validität von Messungen:** Online-Umfragen müssen in ihrem Umfang kurzgehalten werden, um einerseits überhaupt Personen der Zielgruppe für die Teilnahme zu gewinnen und um andererseits einen Dropout während der Befragung zu vermeiden. Daher werden häufig Einzelindikatoren zur Erfassung von Konstrukten wie dem erlebten Stress eingesetzt, um den Fragebogen kurz zu halten. Diese Limitation lässt sich aber nur schwer mit Ansprüchen an die Reliabilität und Validität einer Studie bzw. Umfrage (insbesondere der Erfassung von Konstrukten, für die in der Regel Skalen mit mehreren Items notwendig sind; Stichwort: Messfehler, Inhaltsvalidität) vereinbaren. *Es ist daher ein kritischer Blick auf die Operationalisierung der interessierenden Variablen der jeweiligen Online-Befragungen zu werfen.* Informationen zu Reliabilität und Validität der Messungen sind dabei wichtig.
- 4. Theorie:** Aufgrund der knappen Vorbereitungszeit ist kritisch zu hinterfragen, inwiefern den jeweiligen Online-Umfragen (und den anschließenden Analysen) eine solide theoretische Basis zugrunde gelegt wurde, oder ob sich die Umfragen auf reine deskriptive „Situationsanalysen“ beschränken. Aus statistischer Sicht bedingen insbesondere Signifikanztests theoriegeleitete Hypothesen als Grundlage. *Sofern Ergebnisse auf Basis von Signifikanztests berichtet werden, ist daher kritisch zu prüfen, ob theoriebasierte Hypothesen zugrunde liegen.* Allenfalls besteht die Gefahr des Multiplen Testens (Werden an einem Datensatz mehrere Hypothesen geprüft, wird es zunehmend wahrscheinlicher einzelne statistisch signifikante Ergebnisse zu erzielen). In diesem Zusammenhang ist auch die Frage nach der Verwendung von etwaigen Korrekturverfahren in den Blick zu nehmen.
- 5. Methodik:** Auch wenn manche Online-Umfragen offene Antwortformate einsetzen oder gar um qualitative Interviews ergänzen, fehlt der hier gelieferten Übersicht ein ergänzender Überblick an qualitativen Arbeiten zum Thema. Erste qualitative Studien liegen bereits vor (siehe bspw. das Beiheft 16 der deutschen Schule). Da die Erhebung und Auswertung bei qualitativen Studien meist länger dauern, ist zu erwarten, dass in Zukunft weitere erscheinen werden. *Eine gemeinsame, vergleichende Interpretation von quantitativen und qualitativen Befunden (insbesondere mixed-methods Studien bzw. Umfragen) ist einer Interpretation von monomethodischen Studien bzw. Umfrage vorzuziehen.*
- 6. Multi-Informant Studies:** Die Aussagekraft von Online-Umfragen kann erhöht werden, indem dieselben Items von mehreren Personengruppen (z.B. Schülerinnen und Schülern, Lehrerinnen und Lehrern) aus deren jeweils spezifischen Perspektive eingeschätzt werden. Manche der hier gelisteten Online-Umfragen nutzten die Möglichkeit interessierende Aussagen aus unterschiedlichen Informationsquellen einschätzen zu lassen, um so die Validität der Befunde zu stärken. Insbesondere bei der Analyse von Zusammenhängen auf Basis von Informationen nur einer Personengruppe (z.B. Schülerinnen und Schüler) ist mit dem Common Method Bias, in der Regel einer Überschätzung der Zusammenhänge aufgrund der Erfassung aller Variablen in nur einer Personengruppe, zu rechnen. *Dort wo möglich,*

sollte bei der Interpretation von Befunden daher kritisch geprüft werden, ob derartige Common Method Probleme vorliegen könnten.

- 7. Weitere Einflussfaktoren auf die Validität von Ergebnissen.** Als weitere Einflussfaktoren auf die Qualität von Studienergebnissen ist die Frage von fehlenden Antworten (Non-Response) zu beachten. Der Unit-Non-Response (d.h. eine „gewählte“ Person, hat den gesamten Fragebogen nicht ausgefüllt) kann zu einer Einschränkung der externen Validität führen (siehe oben). Der Item-Non-Response (d.h. wenn Befragte einzelne Fragen nicht beantworten) kann auch zu verzerrten Ergebnissen führen. *Grundsätzlich ist zu prüfen, ob bei einzelnen Fragen hohe Anteile an Item-Non-Response bestehen, ob geprüft wurde, ob es einen systematischen Non-Response gibt und ob etwaige Korrekturverfahren (Gewichtung, Imputation, ...) durchgeführt wurden.* Auch die Verarbeitung der Daten bei Onlineumfragen ist in den Blick zu nehmen, bzw. gilt es der Frage nachzugehen, ob die Verarbeitung transparent beschrieben ist (Welche Fälle wurden nach welchen Kriterien für die Analysen herangezogen?). Wurden etwa Befragte, die aufgrund von „Müdigkeit“ die Befragung abgebrochen haben aus der Ergebnisdarstellung ausgeschlossen. Ist diese Gruppe durch ein spezifisches Antwortverhalten (z.B. hohe Stressbelastung) gekennzeichnet, kann es zu einer Verzerrung der Ergebnisse kommen (vgl. Repräsentativität). *Es ist also zu prüfen, ob die Verarbeitung der Daten (insbesondere der Ein- und Ausschluss von Fällen) ausreichend beschrieben ist.*

**Open Science und Transparenz des Forschungsprozesses.** Für die Beurteilung vieler dieser angeführten Qualitätsmerkmale und Herausforderungen ist die Transparenz des Forschungsprozesses ausschlaggebend. Einerseits mussten/müssen Studien/Umfragen mit knapper Vorbereitungszeit durchgeführt werden. Andererseits mussten/müssen Ergebnisse rasch erzielt werden, um die zu gewinnenden Erkenntnisse zeitnahe nutzen zu können (z.B. bereits für das nächste Schuljahr). Diese Kombination aus zeitlichem Druck und dem Anreiz rascher Erkenntnisse kann die Belastbarkeit von Befunden senken. Vor diesem Hintergrund sollte umso mehr Augenmerk auf Open Science Praktiken gelegt werden, um die Transparenz des Forschungsprozesses zu gewährleisten und schlussendlich Empfehlungen auf belastbare Befunde zu stützen.

Trotz möglicher Limitationen in den bisherigen Online-Umfragen könnte **eine Synthese der berichteten Befunde** ein **valideres Bild** der aktuellen Situation des Lehrens und Lernens während und nach der Schulschließungen aufgrund der Corona-Pandemie liefern. In diesem Sinne hoffen wir, dass der hier geleistete Überblick über bestehende Online-Umfragen zweierlei leisten: Einerseits soll der Überblick Orientierungshilfe für interessierten Leserinnen und Lesern bieten. Andererseits sollen damit die Wissensbasis zum Thema erweitert werden und weitere Forschung in diesem Feld angestoßen werden.

## Literatur

- Biemer, P.P. & Lyberg L.E. (2003). *Introduction to Survey Quality*. New Jersey: John Wiley & Sons, Inc.
- Groves, R. M., & Lyberg, L. (2010). Total survey error: Past, present, and future. *Public opinion quarterly*, 74(5), 849-879.
- Huber, S. G. & Helm, C. (2020). COVID-19 and schooling: evaluation, assessment and accountability in times of crises—reacting quickly to explore key issues for policy, practice and research with the school barometer. *Educational Assessment, Evaluation and Accountability* 32(2), 1-34.  
<https://doi.org/10.1007/s11092-020-09322-y>.