Universität Zürich[UZH]

Institut für Bildungsevaluation
Assoziiertes Institut der Universität Zürich

Texte schreiben 2021 – Standardisierte Erfassung von Schreibkompetenzen

Management Summary

Gerard Adarve, Sarah Smirnow, Benjamin Wolf & Stéphanie Berger

Zürich, 23. März 2022

http://www.ibe.uzh.ch

Ausgangslage

Seit dem Schuljahr 2010/11 wird das computergestützte Testsystem «Stellwerk» im Kanton Zürich in der 8. Klasse flächendeckend eingesetzt. «Stellwerk» umfasst Tests für die reproduktiven Fähigkeiten in den Fachbereichen Deutsch, Englisch, Französisch, Mathematik sowie Natur und Technik. Im Dezember 2010 hat der Bildungsrat des Kantons Zürich beschlossen, das Testsystem «Stellwerk» mit der Erfassung von produktiven Fähigkeiten im Fachbereich Deutsch zu ergänzen und anhand eines standardisierten Verfahrens die Schreibkompetenzen der Schülerinnen und Schüler zu erfassen.

Im Schuljahr 2021/22 wurde der «Stellwerk»-Test zum zehnten Mal mit dem ergänzenden Modul «Texte schreiben» durchgeführt. Das Modul «Texte schreiben» erfasst, inwiefern die Schülerinnen und Schüler Texte verständlich, korrekt und je nach Zielsetzung adressatengerecht formulieren, überzeugend argumentieren oder Sprache ästhetisch ansprechend und kreativ einsetzen.

Durchführung

Die Durchführung des Moduls «Texte schreiben» fand am 15. und 16. November 2021 statt. Die Teilnahme war für alle Schülerinnen und Schüler der 8. Klasse im Kanton Zürich obligatorisch. Insgesamt verfassten 11'445 Schülerinnen und Schüler während 60 Minuten einen Text. Die Schülerinnen und Schüler wählten dabei eines der drei vorgelegten Themen aus:

- Mein Wunsch
- Allein im Gebirge
- Mode und Markenartikel

Die drei Themen wurden je mittels einer kurzen Einleitung und drei zu beantwortenden Fragen vorgestellt. Die Fragen dienten als Leitfaden und halfen bei der Strukturierung der Texte.

Beurteilungsraster

Um die Texte der Schülerinnen und Schüler bewerten zu können, wurde ein standardisiertes Beurteilungsraster, basierend auf wissenschaftlichen Erkenntnissen zu Textanalyserastern[1], eingesetzt. Das Raster umfasst vier Beurteilungsdimensionen:

- Inhalt (Auftragserfüllung und Aussagekraft des Textes)
- Textaufbau (innere und äussere Gliederung des Textes)
- Sprachrichtigkeit (Rechtschreibung und Grammatik)
- Sprachangemessenheit, Schreibstil und Ästhetik

Die Texte der Schülerinnen und Schüler wurden von zehn Beurteilerinnen und Beurteilern (Raterinnen und Ratern) anhand dieser Dimensionen bewertet. Um ein gemeinsames Verständnis der Beurteilungskriterien über die gesamte Bewertungsphase zu gewährleisten, wurden anschliessend an die ausführliche Einarbeitungsphase insgesamt rund 300 zufällig ausgewählte Texte von allen Raterinnen und Ratern beurteilt und miteinander verglichen. Die Übereinstimmung der individuellen Beurteilungen aller Raterinnen und Rater bei diesen Texten

[1] Nussbaumer, M. & Sieber, P. (1994). Texte analysieren mit dem Zürcher Textanalyseraster. In P. Sieber (Hrsg.), *Sprachfähigkeiten – besser als ihr Ruf und nötiger denn je! Ergebnisse aus einem Forschungsprojekt* (S. 141–186). Aarau: Sauerländer.

Becker-Mrotzek, M. & Böttcher, I. (2011). *Schreibkompetenzen entwickeln und beurteilen.* Berlin: Cornelsen.

wurde mittels Kennwert a_{WG} bestimmt.[2] Bei 80 Prozent der Texte war die Übereinstimmung sehr gut ($a_{WG} > .70$). Bei den restlichen 20 Prozent der Texte konnte zudem eine zufriedenstellende Übereinstimmung erreicht werden ($a_{WG} > .50$). Damit können die Beurteilungen der Raterinnen und Rater aus testtheoretischer Perspektive als zuverlässig bezeichnet werden. Das eingesetzte Beurteilungsraster hat sich zudem zur Bestimmung der Schreibkompetenzen der Schülerinnen und Schüler bewährt, indem gute Schülerinnen und Schüler eine hohe Punktzahl erreichten und schwache Schülerinnen und Schüler eine tiefe Punktzahl.

Im Anschluss an die Beurteilungsphase wurden die beurteilten Texte jeweils mit dem angehefteten Beurteilungsraster an die Schülerinnen und Schüler zurückgesandt.

Testergebnisse

Das erste Thema «Mein Wunsch» wurde von 35 Prozent der Schülerinnen und Schüler gewählt. Das zweite Thema «Allein im Gebirge» wurde von 37 Prozent der Schülerinnen und Schüler bearbeitet. 28 Prozent der Schülerinnen und Schüler verfassten ihren Text zum dritten Thema «Mode und Markenartikel». Weniger als 1 Prozent der Texte konnte keinem oder mehreren der drei Themen zugeordnet werden.

Die Testergebnisse der Schülerinnen und Schüler wurden mit Hilfe der Item-Response-Theorie ermittelt. Dabei konnte die verbleibende, leicht unterschiedliche Beurteilungsstrenge der Raterinnen und Rater korrigiert werden, sodass eine faire Beurteilung möglich war. Die mit diesem Verfahren ermittelten Ergebnisse wurden in eine standardisierte Normalverteilung transformiert, welche einen Mittelwert von 500 und eine Standardabweichung von 100 Punkten aufweist. Diese Skala hat die Eigenschaft, dass rund 68 Prozent der Ergebnisse zwischen 400 und 600 Punkten liegen, rund 95 Prozent zwischen 300 und 700 Punkten und nahezu alle Ergebnisse zwischen 200 und 800 Punkten.

Abschliessend kann festgehalten werden, dass die Erfassung der produktiven Kompetenzen im Fachbereich Deutsch anhand des Moduls «Texte schreiben» aus einer testtheoretischen Perspektive als zuverlässig betrachtet werden kann. Wenn standardisierte Beurteilungskriterien sowie eine ausführliche Einarbeitungsphase der Raterinnen und Rater mit einer klaren Definition des gemeinsamen Beurteilungsmassstabs gewährleistet sind, ist die Erfassung von produktiven Kompetenzen der Erfassung von reproduktiven Kompetenzen mittels Leistungstests ebenbürtig. Die Ergebnisrückmeldung aller «Stellwerk»-Tests auf derselben Skala darf dennoch nicht darüber hinwegtäuschen, dass die Ergebnisse des Moduls «Texte schreiben» unabhängig von den anderen Testergebnissen der «Stellwerk»-Tests zu interpretieren sind und sich, im vorliegenden Fall, ausschliesslich auf die 11'445 beteiligten Schülerinnen und Schüler der 8. Klasse im Kanton Zürich beziehen.

[2] LeBreton, J. & Senter, J. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods, 11*, 815–852.

Brown, R. D. & Hauenstein, N. M. A. (2005). Interrater agreement reconsidered: An alternative to the r_{WG} indices. *Organizational Research Methods, 8*, 165–184.