



**Universität
Zürich** ^{UZH}

**Institut für Bildungsevaluation
Assoziiertes Institut der Universität Zürich**

**Texte schreiben 2015 – Standardisierte Erfassung von
Schreibkompetenzen
Management Summary**

Jeannette Oostlander, Caroline Wölfle

Zürich, 3. März 2016

<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 2015/16 wurde der «Stellwerk-Test» zum vierten Mal mit dem ergänzenden Modul «Texte schreiben» durchgeführt. Das Modul «Texte schreiben» hat zum Ziel, zu erfassen, 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 30. November und 1. Dezember 2015 statt. Die Teilnahme war für alle Schülerinnen und Schüler der 8. Klasse im Kanton Zürich obligatorisch. Insgesamt verfassten 10'386 Schülerinnen und Schüler während 60 Minuten einen Text. Die Schülerinnen und Schüler wählten dabei eines der drei vorgelegten Themen:

- Musik hören als Gefahrenquelle
- Aus Spass wird Ernst
- Fotos und Filme im Internet

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 Textanalyserastern¹, eingesetzt. Das Raster umfasst vier Beurteilungsdimensionen:

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

Die Texte der Schülerinnen und Schüler wurden von acht Korrektorinnen und Korrektoren (Ratern) anhand dieser Dimensionen beurteilt. Um ein gemeinsames Verständnis der Beurteilungskriterien über die gesamte Korrekturzeit zu gewährleisten, wurden anschliessend an die ausführliche Einarbeitungsphase insgesamt 317 zufällig ausgewählte Texte von allen Ratern beurteilt und miteinander verglichen. Die Übereinstimmung der individuellen

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

Beurteilung aller Rater bezüglich dieser Texte wird mittels Kennwert a_{WG} bestimmt². Bei 80% der Texte konnte eine zufriedenstellende ($a_{WG} > .50$) oder sogar sehr gute Übereinstimmung ($a_{WG} > .70$) erreicht werden. Damit können die Beurteilungen der 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 erreichen 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 «Musik hören als Gefahrenquelle» sowie das zweite Thema «Aus Spass wird Ernst» wurden je von rund 30 Prozent der Schülerinnen und Schüler gewählt. Etwa 40% der Schülerinnen und Schüler verfassten ihren Text zum dritten Thema «Fotos und Filme im Internet». Die beiden Themen «Musik hören als Gefahrenquelle» und «Aus Spass wird Ernst» wurden von Knaben bevorzugt. Das dritte Thema «Fotos und Filme im Internet» wurde öfter von Mädchen gewählt.

Die Test-Ergebnisse der Schülerinnen und Schüler wurden mit Hilfe der Item-Response-Theorie ermittelt. Dabei konnte die verbleibende, leicht unterschiedliche Beurteilungsstrengere der Rater korrigiert werden, sodass eine faire Beurteilung möglich war. Die mit diesem Verfahren bestimmten Ergebnisse wurden in eine standardisierte Normalverteilung transformiert, welche analog zur «Stellwerk»-Skala 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.

Die Mädchen erreichen im Durchschnitt 517 Punkte auf der Stellwerkskala, die durchschnittliche Punktzahl der Knaben liegt bei 487. Diese Differenz in der Fähigkeit von 30 Punkten ist statistisch signifikant ($p < .001$), jedoch von geringer Bedeutsamkeit (Effektstärke $d = .30$). Eine Aufteilung der Ergebnisse nach Schultypen zeigt nach wie vor, wie sinnvoll eine schultypen-unabhängige Beurteilung sein kann. Beispielsweise schreiben rund 10 Prozent der Schülerinnen und Schüler der Sek B Texte, welche in ihrer Beurteilung über dem Mittelwert der Sek A liegen.

Abschliessend kann gesagt werden, dass die Erfassung der produktiven Kompetenzen im Fachbereich Deutsch 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 Beurteilungsmasstabs eingehalten werden, ist die Erfassung von produktiven Kompetenzen der Erfassung von reproduktiven Kompetenzen mittels Leistungstests ebenbürtig. Die Ergebnismrückmeldung 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 10'386 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.