Mit welchen Tricks Lehrpersonen vertuschen, dass Prüfungen nicht objektiv sind

Eine Prüfung ist objektiv, wenn unterschiedliche Prüfer*innen zum selben Ergebnis kommen, wenn sie die Ergebnisse beurteilen. Nimmt man lediglich Prüfungsaufgaben und Ergebnisse von Schüler*innen, dann kommen Prüfer*innen zu krass unterschiedlichen Ergebnissen. Nehmen wir eine Mathematik-Aufgabe:

Wie viele Quadratzahlen gibt es, die kleiner sind als 10^6?

Betrachten wir folgende Antworten:

- $10^3 - 1$
- 10^3
- $\sqrt{1'000'000}$
- »10^6 ist eine Million, das ist das Quadrat von Tausend. Ich denke, es gibt weniger als Tausend.«

Welche Antwort gibt wie viele Punkte? Sobald wir das diskutieren, merken wir, dass wir klären müssen, ob 0 eine Quadratzahl ist. Und wir müssen klären, ob Wurzeln stehen gelassen werden dürfen und ob Antworten in natürlicher Sprache formuliert sein dürfen, ob sie vage sein dürfen.

Alle Antworten weisen gewisse Qualitäten auf, sie zeugen von Kompetenzen. Könnten Prüfungen diese Kompetenzen messen, dann wäre klar, wie sie einzuschätzen sind. Da Prüfungen bzw. Prüfungsaufgaben Kompetenzen nicht messen, würden Mathematiklehrpersonen zu sehr unterschiedlichen Punktzahlen für diese Lösungen kommen. Einige würden einzelnen Lösungen 0 Punkte geben, andere würden allen die volle Punktzahl geben. Einige würden bestimmte Teilpunkte vergeben, andere andere.

Das ist bei allen Aufgaben so. Sobald eine andere Lehrperson korrigiert, kommt sie zu einer anderen Einschätzung. Nun könnte man das als Beleg auffassen, dass Prüfungsaufgaben nicht objektiv sind. Wenn man ehrlich wäre, müsste man das als Beleg dafür auffassen.

In den Erziehungswissenschaften passiert aber was anderes, was der Thread des Erziehungswissenschaftlers Kris Besa (unten) zeigt. Die fehlende Objektivität wird mit Tricks kaschiert. Die Tricks funktionieren so:

1. Willkürliche Vorgaben bei der Prüfung schränken die Antwortmöglichkeiten ein. Zum Beispiel wird vorausgesetzt, dass Schüler*innen eine bestimmte

Definition von Quadratzahl anwenden, die sie im Unterricht gelernt haben (egal, wie problematisch diese Definition sein mag). Oder es wird klar gemacht, dass Ergebnisse Zahlen sein müssen, die doppelt unterstrichen werden etc.

2. Die Aufgabe wird in Teilschritte zerlegt, für die Teilpunkte vergeben werden. Diese Aufteilung oder Zerlegung ist willkürlich, sie orientiert sich nicht an tatsächlichen Kompetenzen, sondern modelliert Kompetenzen willkürlich und belohnt dann bestimmte Fertigkeiten, andere aber nicht. (Es wäre auch eine Kompetenz, bei einer solchen Aufgabe zu beschreiben, weshalb man nicht auf die Lösung kommt.)
3. Mit Erwartungshorizonten oder Musterlösungen wird vorgegeben, wie Punkte verteilt werden dürfen.
4. Die Antworten aller Schüler*innen werden verglichen, danach entscheidet eine Lehrperson, welche Lösungen zu wie vielen Punkten führen (sie belohnt dann z.B. Überlegungen, die nur eine Schülerin gemacht hat, nicht – vergibt aber Punkte für fehlerhafte Ansätze, die viele Schüler*innen gewählt haben).

Argumentativ kann dann gesagt werden, Prüfungen seien nur dann objektiv (oder könnten nur dann messen), wenn sie diese Vorgaben machen, an eine soziale Gruppe angepasst werden und mit Erwartungshorizonten verstehen werden. Damit wird aber kaschiert, dass sie halt weder objektiv sind noch messen, sondern dass man einfach diesen Eindruck vermeiden will.

Um ein Bild zu verwenden: Das wäre, als würde man einen Maßstab nehmen und ihn verwenden, um eine Strecke abzumessen. Manchmal ist die gemessene Strecke 12cm lang, manchmal ist sie 9cm, manchmal 10cm. Danach würde man sagen: Wir nehmen nur die Ergebnisse, die gleich sind (Erwartungshorizont). Das ist die richtige Messung. Dann scheint der Maßstab genau zu sein, obwohl er es nicht ist.

(Historisch waren Prüfungen und Noten nicht als objektive Messung konzipiert, bevor diese positivistischen Vorstellungen im 19. Jahrhundert an Bedeutung gewonnen haben. Man hat also eine bestehende Praxis genommen und so getan, als handle es sich um eine Messung, obwohl das nie so gedacht war.)

Die toxische Rangordnung: Warum Noten Über- und Unterlegenheit erzeugen

In einem Podcast erzählt Maria Bamford, wie sie als Patientin in einer Klinik für Essstörungen von anderen Patientinnen bewertet und beurteilt wurde: Es gebe, so ihre Beobachtung, Hierarchien, wer eine gravierende und belastendere Essstörung habe. Das mag leicht übertrieben sein, zeigt aber, dass Menschen alles benutzen können, um anderen Über- und Unterlegenheit vorzuführen, um eine Hackordnung zu etablieren.

Das ist toxisch: Wir schreiben uns einen Wert zu, indem wir uns über andere erheben. Wir sprechen uns einen Wert ab, indem wir andere als überlegen wahrnehmen. Das kann man auf Gesundheit, auf Aussehen, auf Einkommen, auf Familie oder auf den Beruf beziehen: Sobald eine Bewertung stattfindet, erzeugt sie Scham und negative Gefühle. Deshalb ist toxisch der richtige Begriff dafür, die Bewertungen vergiften etwas, was nicht giftig sein sollte. Ob jemand viele Kinder, ein Kind, keine Kinder hat – alle Personen leben ein Leben, das eine Geschichte und eine Bedeutung hat. Hier eine Wertung einzuführen, ist *nur* problematisch, es ergibt keinen Sinn. Genauso ist es mit Berufen: Was auch immer eine Person tut, bedeutet für sie und andere etwas. Einige Berufe über andere zu stellen, das schadet nur.

Damit sind wir bei den Noten: Kinder lernen, dass und wie sie sich über andere stellen können (oder wie und dass sie sich anderen unterordnen müssen). Leider lernen sie das auch über Schulnoten. Kinder lernen in der Schule. Sie strengen sich an, machen Fehler, sind manchmal eifrig dabei und manchmal gelangweilt oder wenig motiviert. Wenn nun mit Prüfungen und Noten eine Bewertung darüber gelegt wird, die einen Vergleich erlaubt, entsteht sofort die Frage, wer in Mathe »gut« ist und wer »schlecht« ist. Mathe kann plötzlich zu einem Wert werden, der einer Person und ihrer Arbeit zugeordnet ist. Einige sind »besser« als andere. Was passiert, sobald eine Klasse Note erhält, ist verheerend: Sie werden nicht auf Prüfungsleistungen bezogen, sondern auf die ganze Person. »Du bist schlecht in Mathe« beginnt mit »Du bist schlecht.« Noten verunmöglichen, dass alle gut sein können. Die Aufwertung der einen wird mit der Abwertung anderer verkauft.

Auch wenn es eine faire und sinnvolle Form gäbe, Mathe-Leistungen zu vergleichen (was nicht der Fall ist), wäre es problematisch, diese Bewertungen so einzuführen. Viele Menschen rechtfertigen sie damit, dass Kinder sich mit anderen vergleichen wollen. Das tun sie, weil sie gelernt haben, daraus einen Wert abzuleiten. Könnten wir unabhängig von Beurteilungen sicher sein, dass unsere Anstrengungen etwas bedeuten, bräuchten wir nicht die Sicherheit, etwas besser als andere zu können.

Aus der toxischen Rangordnung entsteht der Teufelskreis der Meritokratie: Wer auf der Rangliste weiter oben steht, muss Privilegien erhalten. Damit diese Privilegien

verteilt werden können, braucht es eine Rangliste. Verkürzt heißt das dann: »Leistung muss sich lohnen«. Nein, muss sie nicht. Schulische Selektion braucht es nämlich in dieser Argumentation, damit sich Leistung lohnt. Und Leistung lohnt sich, weil es schulische Selektion gibt. Das gilt auch auf den Verweis auf andere ungerechte Formen von Meritokratie: Auch ungleiche Löhne oder Lebensverhältnisse sind nicht verdient, sondern entstehen aus einem System, das Menschen bewertet und einander über- oder unterordnet.

Menschen strengen sich an, wenn sie dafür gute Bedingungen vorfinden. Es braucht keine künstlichen Formen von Beurteilung, keine Abwertung von anderen. Es gibt keine Form von Bewertung, welche nicht eine Rangordnung erzeugt und so instrumentalisiert wird, um Menschen abzuwerten. Deshalb hat Bewertung in einer Schule nichts verloren. Sie ist toxisch. Feedback ist das nicht – es bezieht sich auf Menschen und schafft keine künstlichen Ordnungen und Machtverhältnisse.