

## **Müssen Prüfungen wirklich alles können? (ein Beitrag zur Diskussion von Andreas Stöhr)**

Nach dem BBiG soll mit Prüfungen die berufliche Handlungsfähigkeit festgestellt werden - die Prüflinge müssen nachweisen, dass sie über die erforderlichen Fertigkeiten, Kenntnisse und Fähigkeiten verfügen, um künftig in ihrem Beruf selbstständig arbeiten zu können. Dabei müssen Prüfungen geradezu unglaubliches leisten: Sie sollen authentisch sein, also realitätsnah und auf den jeweils erlernten Beruf bezogen. Dann soll eine Prüfung noch objektiv, reliabel und valide sein und darüber hinaus auch noch zuverlässig zwischen guten und weniger guten Prüflingen unterscheiden, und - nicht zuletzt - möglichst nicht viel kosten. Prüfungen müssen einer ganzen Reihe von Ansprüchen genügen. Diese Ansprüche schließen sich zwar gegenseitig nicht unbedingt aus, behindern sich aber. Es stellt sich die grundsätzliche Frage, ob Prüfungen überhaupt allen diesen Anforderungen (gleichzeitig) gerecht werden können.

Berufliche Handlungsfähigkeit kann in den Kompetenzdimensionen Fach-, Methoden- und Sozialkompetenz beschrieben werden. Fach- und Methodenkompetenz lassen sich mit der Prüfung relativ konkret nachweisen. Etwas schwieriger wird es bei der Sozialkompetenz, da gezeigtes Verhalten nicht immer so klar definierbar ist, wie zum Beispiel technische Daten. Sozialkompetenz kann immer nur bezogen auf die berufliche Anforderung (z.B. im Beratungsgespräch) beobachtet werden. Dann lassen sich Einstellungen und Werte einer Person aus dem gezeigten Verhalten ableiten. Welche Einstellungen und Werte eine Person außerhalb der Prüfungssituation tatsächlich hat, das kann ein Prüfungsausschuss nur vermuten.

Prüfungen sollen authentisch und an Prozessen orientiert sein, sie sollen also möglichst realen beruflichen Situationen entsprechen und an Arbeits-, Produktions- und/oder Geschäftsprozessen orientiert sein. Geeignet wäre daher eine Prüfung, die einen solchen Prozess vollständig abbildet oder simuliert. In diesem Rahmen ließen sich auch soziale und persönliche Kompetenzen (fachbezogen!) problemlos messen. Könnte so ein Prozess nicht nur als Simulation, sondern auch noch in der betrieblichen Arbeitsrealität beobachtet werden, wäre man der Abbildung beruflicher Handlungsfähigkeit (oder beruflicher Handlungskompetenz) sehr nahe. Leider sind hier (zeitliche und personelle) Grenzen gesetzt, die durch den Prüfungsaufwand bestimmt werden.

Die Gütekriterien Objektivität, Reliabilität und Validität werden immer wieder genannt, wenn es um die Qualität von Prüfungen und Prüfungsaufgaben geht.

Objektivität bedeutet, dass externe Faktoren die Prüfung nicht beeinträchtigen dürfen und dass verschiedene Prüfer – unabhängig voneinander – das Prüfungsergebnis in gleicher Weise beurteilen bzw. zur gleichen Ergebnisbewertung kommen. Reliabilität bedeutet, dass die Prüfung das, was sie misst, auch möglichst genau misst. Hier geht es um die Messgenauigkeit des Instruments Prüfung. Validität bedeutet, dass die Prüfung das misst, was sie messen soll, nämlich berufliche Handlungsfähigkeit.

Die Ansprüche aus diesen Gütekriterien klingen zunächst trivial, sind aber in der Praxis nicht leicht zu erfüllen. Oft kommen verschiedenen Beurteiler, die ein und dieselbe Arbeit beurteilen sollen, zu recht unterschiedlichen Ergebnissen. Die Messgenauigkeit, insbesondere bei gezeigtem (menschlichen) Verhalten, ist nicht immer einfach zu bestimmen, da Übergänge im Verhalten nicht immer leicht zu erkennen und daher auch nicht leicht zu skalieren sind (zum Beispiel in einem Verkaufsgespräch, wo der Übergang von „hoch verkaufsmotiviert“ zu „(feindlich) aggressiv“ fließend sein kann). Und ob eine Prüfung immer das misst, was sie messen soll, darf - ab und zu - in Frage gestellt werden.

Es stellt sich die grundsätzliche Frage, ob Prüfungen tatsächlich immer das messen, was sie auch messen sollen (nämlich berufliche Handlungsfähigkeit). Um eine Antwort auf diese Frage zu finden, können Prüfungen (zum Beispiel) testtheoretisch betrachtet werden. Die Prüfung selbst kann dabei als Testbatterie<sup>1</sup>, die aus mehreren Einzeltest besteht, angesehen werden. Bei solcher Betrachtung liegt es nahe, die Prüfung mit den Kriterien der klassischen<sup>2</sup> oder der probabilistischen<sup>3</sup> Testtheorie zu beurteilen. Es stellt sich jedoch die Frage, ob eine rein testtheoretische bzw. diagnostische Betrachtung von Prüfungen ausreichend ist, um berufliche Handlungsfähigkeit hinreichend erfassen zu können. Lienert & Raatz (1998) definieren den Begriff „Test“ wie folgt:

*„Ein Test ist ein wissenschaftliches Routineverfahren zur Untersuchung eines oder mehrerer empirisch abgrenzbarer Persönlichkeitsmerkmale mit dem Ziel einer möglichst quantitativen*

<sup>1</sup> In der diagnostischen Psychologie besteht eine Testbatterie aus unterschiedlichen Testverfahren, die zur Beantwortung ein und derselben Frage eingesetzt werden.

<sup>2</sup> Nach der klassischen Testtheorie besteht jede Messung aus einem (wahren) Merkmalsanteil und einem (unsystematischen) Fehleranteil, wobei der Fehleranteil möglichst klein gehalten werden soll.

<sup>3</sup> Mit Methoden der probabilistischen Testtheorie wird von gemessenen Daten (aus einer Untersuchung) auf eine diesen Daten zugrunde liegende (latente) Variable geschlossen.

*Aussage über den relativen Grad der individuellen Merkmalsausprägung.“ (Lienert& Raatz 1998, S. 3)*

Auf Nachfrage würden sicherlich nicht viele Prüferinnen und Prüfer ihre Prüfungen als „wissenschaftliche Routineverfahren“ bezeichnen. Solche wissenschaftlichen Routineverfahren könnten natürlich Bestandteil einer beruflichen Abschlussprüfung sein, würden aber zur Beurteilung beruflicher Handlungsfähigkeit kaum ausreichen. Tatsächlich gibt es auch einen alternativen Ansatz. Dieser besteht in einer Erweiterung des (rein) diagnostischen Ansatzes um konzeptionelle Qualitätskriterien, die sich aus dem Konzept der beruflichen Handlungskompetenz ergeben:

*„Während bei den diagnostischen Gütekriterien die Genauigkeit der Messungen im Vordergrund steht, verlangen die konzeptionellen Qualitätskriterien, dass vor allem auch schwer messbare berufliche Anforderungen bzw. Qualifikationen geprüft werden. Bei der Erstellung von Prüfungsaufgaben und der Beurteilung ihrer Lösungen sind deshalb beide Arten von Qualitätskriterien in Einklang zu bringen.“ (Ver.di-Prüferhandbuch, S. 43).*

Statt also nur Messgenauigkeit – als alleinigen Maßstab – für die Qualität von Prüfungen zu nehmen, muss diese mindestens um die Qualitätskriterien Handlungsorientierung und Praxisnähe erweitert werden. Die Hinzuziehung weiterer Qualitätskriterien schafft – im Hinblick auf die klassischen Qualitätskriterien (Validität, Reliabilität, Objektivität) – eine gewisse Unschärfe, lässt sich doch die Messgenauigkeit von Prüfungsaufgaben nicht mehr so einfach bestimmen, unter anderem schon deshalb nicht, weil in der betrieblichen Praxis oft komplexe und nicht standardisierte Arbeitsaufgaben vorliegen.

Es gilt zu bedenken, dass die Abbildung beruflicher Handlungsfähigkeit das Ziel der Prüfung ist und nicht die graduelle Unterscheidung der Qualität von Lösungsansätzen - von denen es in der Praxis ohnehin meistens mehrere gibt - oder der Ausprägungsgrad von Persönlichkeitsmerkmalen bestimmter Personen.

Eine Prüfung nach den Regeln der probabilistischen Testtheorie ermöglicht natürlich hinreichende quantifizierbare Unterscheidungsmöglichkeiten zwischen den Ausprägungsgraden von Merkmalen. Es ließe sich zum Beispiel damit feststellen, dass Auszubildender A eine gestellte Aufgabe besser gelöst hat als Auszubildender B und dieser wiederum besser als Auszubildender C usw. Diese Unterscheidung wird zwar von Betrieben, Berufsschulen und nicht zuletzt von den Auszubildenden selbst gewünscht, ist aber nur mit bestimmten standardisierten Aufgabentypen einwandfrei zu bewirken. Berufliche

Handlungsfähigkeit und Praxisnähe lassen sich mit solchen Aufgaben nur bedingt erzeugen. Die fachbezogene Feststellung und Beurteilung sozialer Kompetenzen zum Beispiel ist mit diesem standardisierten Aufgabentypus nicht oder nur schwerlich möglich.

Es bleibt also ein Dilemma: Standardisierte (wo möglich computergestützte) Prüfungen haben einen hohen Grad an Messgenauigkeit und Vergleichbarkeit, lassen es aber oft an Handlungsorientierung und Praxisnähe fehlen. Prüfungen, die Produktions- und Geschäftsprozesse handlungsorientiert und praxisnah abbilden, fehlt es oft an hinreichender Vergleichbarkeit und Messgenauigkeit. Beide Prüfungstypen haben ihre Stärken und Schwächen, hinzu kommt auch noch die Frage nach den Kosten einer Prüfung. Da sich dieses Problem nicht einfach lösen lässt, finden die Diskussionen zum Thema „Messgenauigkeit vs. Authentizität“ auch kein Ende.

Um nun die obige Frage „Müssen Prüfungen wirklich alles können?“ beantworten zu können, braucht es an dieser Stelle noch eine kurze Zusammenfassung:

Nicht jede Kompetenzdimension lässt sich mit einer Prüfung hinreichend abbilden und bewerten. Insbesondere persönliche Einstellungen und Werte lassen sich – im Sinne von sozial erwünschtem Verhalten – für die Prüfungsdauer auch gut vortäuschen.

Die Authentizität von Prüfungen ist überaus erstrebenswert, lässt sich aber aus zeitlichen und personellen Gründen nicht beliebig herstellen.

Ob Prüfungen tatsächlich immer dem Kanon der (klassischen) Gütekriterien gerecht werden darf bezweifelt werden, sie müssen es aber auch nicht. Der moderne Arbeitsalltag ist eher von komplexen Situationen geprägt. Ein standardisiertes Vorgehen ist - zum Beispiel in der industriellen Produktion - sicherlich noch häufig notwendig, aber längst haben individuelle Kundenwünsche und neue Formen der Betriebsorganisation den Rahmen der normierten Massenproduktion gesprengt.

Der scheinbare Konflikt „Messgenauigkeit vs. Authentizität“ ist künstlich und folgt dem Diktat der „reinen Lehre“. Die Lösung lautet „Messgenauigkeit und Authentizität“ (im Rahmen des Möglichen).

Es empfiehlt sich also, einen pragmatischen Ansatz zu wählen und in jeder Prüfung einen Mix von Prüfungsinstrumenten einzusetzen, der sowohl standardisierte als auch praxisnahe Instrumente und Ansätze umfasst. So können die Schwächen oder Nachteile der jeweiligen Methoden zwar nicht verhindert aber wenigsten – im Hinblick auf das Ziel der Prüfung – reduziert werden.

Literatur:

Gulliksen H.: Theory of Mental Tests. London 1987.

Lienert, G. A. & Ratz, U.: Testaufbau und Testanalyse. Weinheim 1998.

Lienert, G. A.: Schulnoten-Evaluation. Frankfurt am Main 1987.

Lord, F. M. & Novick, M. R.: Statistical Theories of Mental Test Scores. Massachusetts 1968.

Rasch, G.: Probabilistic models for some intelligence and attainment tests. Chicago 1980.

Ver.di – Vereinte Dienstleistungsgewerkschaft (Hrsg.): Das Prüferhandbuch. Eine Handreichung zur Prüfungspraxis in der beruflichen Bildung. Hamburg 2008.