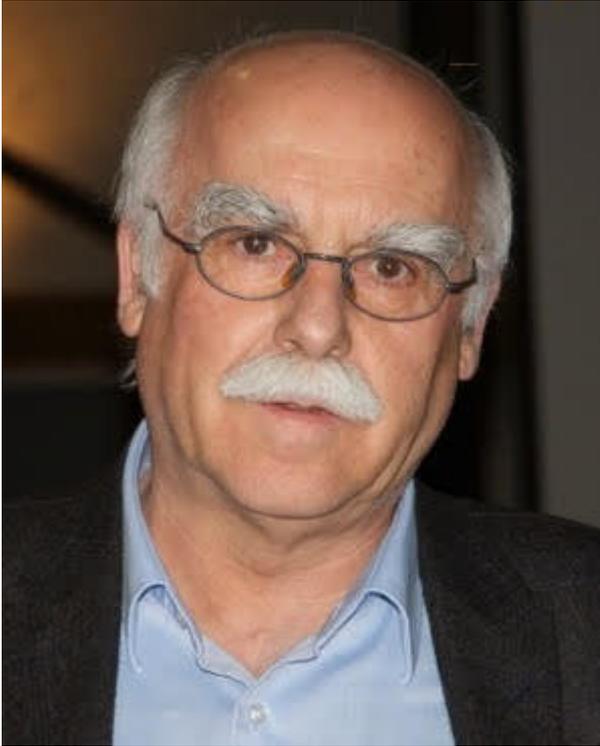


Die PISA-Studien bieten der pädagogischen Praxis wenig Nutzen

Category: Blog

geschrieben von Walter Herzog | 5. Dezember 2019



Es verhält sich mit der Bildung nicht unähnlich wie mit der Zeit. Niemand wird bestreiten, dass wir die Zeit messen können. Aber ebenso wird niemand behaupten, dass die metrische Zeit das Phänomen Zeit vollständig abdeckt. Unsere Uhren wissen nichts vom Leiden an der Zeit, vom Gefühl der Zeitlosigkeit, vom Verweilen im Augenblick oder von der Langeweile. Sie wissen auch nichts von der qualitativen Unterscheidung in Vergangenheit, Gegenwart und Zukunft. Wir mögen noch so lange auf unsere Uhren blicken, wir werden nie erfahren, dass die vergangene Zeit etwas ganz anderes ist als die zukünftige Zeit. Was eine Uhr misst, ist daher nie die ganze Zeit. Etwas Ähnliches trifft auf die Bildung zu. Auch wenn sich Bildung messen lässt, heisst dies nicht, dass die gemessene Bildung uneingeschränkt dem entspricht, was wir unter Bildung verstehen.

Auch wenn sich Bildung messen lässt, heisst dies nicht, dass die gemessene Bildung uneingeschränkt dem entspricht, was wir unter Bildung verstehen.

Entgrenzung von Raum und Zeit

Was tun wir überhaupt, wenn wir etwas messen? Im Falle der Zeit fällt uns eine Antwort nicht allzu schwer. Wir messen die Zeit, indem wir einen regelmässigen natürlichen oder künstlichen Vorgang als Standard setzen, um einen anderen Vorgang damit zu vergleichen. Ob Sonnenuhr, Wasseruhr, Sanduhr, Räderuhr, Pendeluhr oder Quarzuhr, immer gibt es einen sich periodisch wiederholenden Geschehensablauf, der als Massstab dient, um einen anderen Geschehensablauf zu messen.



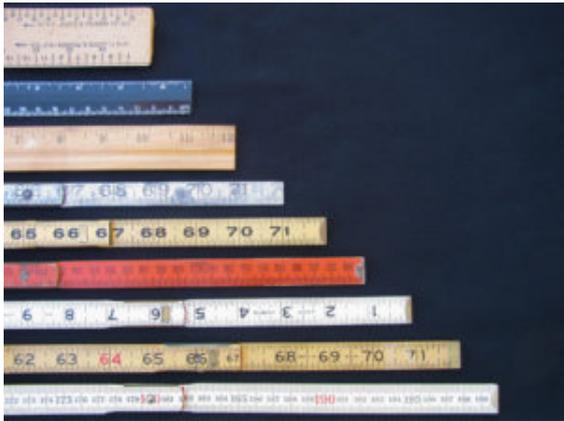
Ob Sonnenuhr, Wasseruhr, Sanduhr, Räderuhr, Pendeluhr oder Quarzuhr, immer gibt es einen sich periodisch wiederholenden Geschehensablauf, der als Massstab dient.

Die Vorstellung, dass Uhren die Zeit *anzeigen* oder *registrieren*, ist daher missverständlich, da sie suggeriert, Zeit sei uns objektiv gegeben und brauche nur verlässlich abgebildet zu werden. Doch eine Uhr bildet die Zeit nicht ab, sondern auferlegt unserer Zeiterfahrung eine bestimmte Ordnung, die es ermöglicht, uns über die Zeit zu verständigen. Wie gut begründet der Vorgang der Zeitmessung auch immer sein mag, es gibt keine Zeit *an sich*. Was Zeit *ist*, legen wir durch die Messung von Zeit überhaupt erst fest.

Wir legen fest, was Bildung ist

Mit der Bildung verhält es sich nicht anders. Auch Bildung gibt es nicht an sich, sondern wir legen fest, was wir unter Bildung verstehen, indem wir uns auf eine lange Tradition des Denkens über Bildung beziehen. Das Gepräge, das Wilhelm

von Humboldt und Hegel dem Bildungsbegriff gegeben haben, ist in vieler Hinsicht auch heute noch relevant, wenn wir über Bildung sprechen. Allerdings war die Messung von Bildung weder für Humboldt noch für Hegel ein Thema. Ihre Auseinandersetzung mit dem Bildungsbegriff erfolgte im Medium der Sprache, deren Nachteil darin liegt, dass wir weniger präzise kommunizieren können, was wir unter Bildung verstehen, als wenn uns ein Instrument zur Verfügung steht, mit dem sich Bildung messen lässt.



Die PISA-Studien stehen in einer Tradition der Standardisierung von Masssystemen.

Die höhere Präzision, die mit einem Messvorgang verbunden ist, hat wesentlich damit zu tun, dass eine Messung von lokalen und personalen Einflüssen weitgehend unabhängig ist. Während die Zeit noch in der ersten Hälfte des 19. Jahrhunderts von den natürlichen Tages- und Jahresrhythmen einer Agrargesellschaft bestimmt wurde und damit *lokale* Zeit war, führte die Industrialisierung zu einer Vereinheitlichung der Zeitmessung. Die aufkommenden Nationalstaaten beförderten die Entwicklung, da sie aufgrund ihrer Verwaltungsbedürfnisse generell an einheitlichen Masssystemen interessiert waren. Offenbar soll nun Gleiches auch mit der Bildung geschehen. Die PISA-Studien stehen in einer Tradition der Standardisierung von Masssystemen, die im 19. Jahrhundert begonnen hat und seither immer mehr Lebensbereiche erfasst. Analog zur metrischen Zeiterfassung soll die Beurteilung von Bildung unabhängig von lokalen Einflüssen sein, da sich nur so über Raum und Zeit hinweg Vergleiche anstellen lassen.

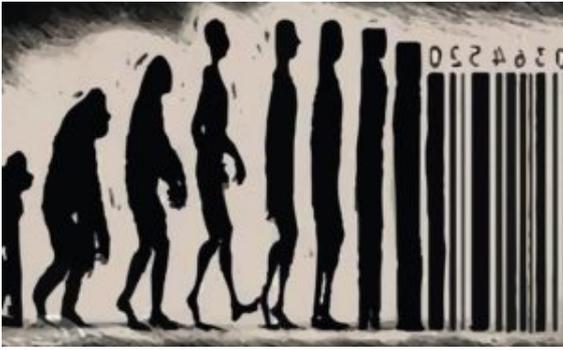
Verschiedene Messniveaus



Was messen wir überhaupt? Bild: AdobeStock

Was aber verstehen wir überhaupt unter einer Messung? Nach einer weit verbreiteten Definition beruht eine Messung auf der Zuordnung von Zahlen zu Eigenschaften von Objekten oder Ereignissen entsprechend einer bestimmten Regel. So können wir Klassen von Objekten mit einer gewissen Eigenschaft (z.B. Menschen mit der Eigenschaft, männlich oder weiblich zu sein) Zahlen zuordnen (den Männern eine 1, den Frauen eine 2 oder umgekehrt) und haben damit einen Messvorgang vollzogen. Allerdings einen äusserst primitiven, da die Zahlen lediglich eine *nominelle* Zuordnung, d.h. eine Benennung, erlauben. Man spricht deshalb von einer Nominalskala.

Etwas gehaltvoller ist die Zuordnung von Zahlen zu Objekten, wenn wir eine Rangreihe bilden können - z. B. die Reihenfolge, in der die Teilnehmer an einem Hundertmeterlauf im Ziel eintreffen. Wenn wir als Zahlensystem die natürlichen Zahlen nehmen und diese beginnend mit 1 vergeben, dann definieren die Zahlen eine Abfolge. Die Zahlen sind nicht mehr bloss Namen, sondern legen eine Ordnung fest, nämlich wer der Schnellste, wer der Zweitschnellste, wer der Drittschnellste etc. ist. Aus der Ordnung als solcher lässt sich aber nicht schliessen, wer der Schnellste *schlechthin* ist, da wir nur die Reihenfolge beim Eintreffen am Ziel registriert und nicht die Laufzeit gestoppt haben. Rang- bzw. Ordinalskalen erlauben eine *qualitative* Ordnung, also gerade nicht das, was wir unter einer Messung verstehen.



Messungen auf dem Niveau von Ratioskalen sind in den Sozialwissenschaften allerdings nicht möglich.

Erst eine Intervallskala, deren Zahlen nicht nur geordnet sind, sondern gleiche Abstände aufweisen und damit *Größenunterschiede* erfassen, erfüllt die Erwartungen an eine Messung. Was einer Intervallskala jedoch noch fehlt, um die Qualität eines physikalischen Masssystems zu erreichen, ist ein definierter Nullpunkt. Ist ein solcher Nullpunkt gegeben wie im Falle der Längen- oder Zeitmessung haben wir es mit einer Ratioskala zu tun, die eine absolute Messung erlaubt. Der Messvorgang wird damit gänzlich unabhängig von lokalen Bedingungen, wie die Zeitmessung bei einem Hundertmeterrennen zeigen kann. Egal wo oder wann die hundert Meter gemessen werden, sie lassen sich über Raum und Zeit hinweg miteinander vergleichen, was es möglich macht, von einem Rekord bzw. «Weltrekord» zu sprechen. Messungen auf dem Niveau von Ratioskalen sind in den Sozialwissenschaften allerdings nicht möglich. Das höchste erreichbare Niveau sind Intervallskalen, wobei auch hier gelegentlich bezweifelt wird, dass dieses Niveau tatsächlich erreicht wird. Besonders schwierig sind daher Vergleiche über die Zeit. Obwohl mit den PISA-Studien solche Vergleiche angestellt werden, sind sie mit Vorsicht zu geniessen.

Tests als Messinstrumente

Wenn Messung darauf beruht, dass ein Massstab definiert wird, mit dem das zu messende Phänomen verglichen wird, dann stellt sich im Falle der Messung von Bildung die Frage, was eigentlich gemessen wird. Was wird womit verglichen? Bei den PISA-Studien stützt man sich auf das Leistungsverhalten der



Schülerinnen und Schüler, denen standardisierte Testaufgaben vorgelegt werden, die sie zu bearbeiten haben. Wenn dann von Wissen, Können oder Kompetenzen gesprochen wird, dann muss man sich aber bewusst sein, dass dies gerade *nicht* gemessen wird. Eine Messung von Wissen oder Kompetenzen ist im strengen Sinn nicht möglich, da psychische Phänomene im Unterschied zu physischen nicht oder nur teilweise *direkt* zugänglich sind. Psychologische Konstrukte wie Persönlichkeit, Gedächtnis, Intelligenz, Problemlösefähigkeit, Kompetenz oder Einstellung, aber auch Schulleistung, sind uns nicht phänomenal gegeben. Vielmehr sind es Konstrukte, deren Realitätsstatus oft unklar ist. Was uns gegeben ist, sind Verhaltensweisen oder Ergebnisse von Verhaltensweisen, die wir als Indikatoren nutzen, um das interessierende Konstrukt zu *erschliessen*.

Psychometrische Tests sind nichts anderes als methodische Hilfsmittel, um individuelle Verhaltensweisen unter kontrollierten Bedingungen auszulösen.

46. Beiliebtheit / Verträglichkeit: Es wird von anderen Kindern ... zum Mitspielen aufgefordert			
sehr oft, regelmäßig	von einzelnen	nur in bestimmten Situationen	nie
<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>
47. Zusammenarbeit: Es stört andere Kinder, die spielen, basteln, zeichnen			
praktisch nie	selten	ab und zu	häufig
<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>
48. Verantwortung: In einer Gruppe übernimmt es ... Verantwortung			
beständig	manchmal	kaum	nur in bestimmten Situationen
<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>
49. Hilfsbereitschaft: Wenn ein anderes Kind Schwierigkeiten hat (mit Spielgegenstand, Ankleiden) ...			
hilft es sofort und ungeduldet	hilft es in bestimmten Situationen und bei bestimmten Kindern	hilft es nur nach Aufforderung	reagiert es gar nicht
<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>
50. Konfliktfähigkeit: Wenn ihm ein anderes Kind etwas weggenommen ...			
wartet es sich durch (oder macht Lösungsvorschläge)	versucht es zuerst, sich selbst durchzusetzen	sucht es sehr nach die Hilfe der Lehrperson	wartet es sich in der Regel nicht / beginnt zu weinen
<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>
51. Selbstkontrolle: Bei gemeinsamen Aktivitäten im Kreis hat es ... Mühe zu warten, bis es an der Reihe ist			
keine (sehr kontrolliert, zurückhaltend)	kaum	noch ab und zu	welche (spricht häufig davon)
<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>

Beurteilungsbogen der Kindergärten in St. Gallen Seite 9

von insgesamt 12 Seiten

Psychometrische Tests sind nichts anderes als methodische Hilfsmittel, um individuelle Verhaltensweisen unter kontrollierten Bedingungen auszulösen. Tests werden so konstruiert, dass sie einem Messverfahren *nahekommen*. Die üblichen Testkriterien - Objektivität, Reliabilität und Validität - sind Ersatzkriterien, um den Mangel an Messqualität zu kompensieren. Die Kriterien sind darauf ausgerichtet, die Durchführung eines Tests sowie die Auswertung der Testergebnisse zu normieren, womit der Anspruch einer Messung, nämlich von lokalen und situativen Bedingungen unabhängig zu sein, wenigstens annäherungsweise eingelöst werden kann. Dies bringt das folgende Zitat von Lee Cronbach aus seinen *Essentials of Psychological Testing* pointiert zum Ausdruck: «A standardized test is one in which the procedure, apparatus, and scoring have been fixed so that precisely the same testing procedures can be followed *at different times and places*» (S. 27 - Hervorhebung W. H.). Tests wie sie im Rahmen von PISA und vergleichbaren Schulleistungsstudien zum Einsatz kommen erfüllen damit eine analoge Funktion wie Uhren im Falle der Zeiterfassung, nämlich Bildung unabhängig von lokalen Bedingungen zu messen. Die Qualität der Messung von Schülerleistungen ist allerdings bedeutend geringer als die Qualität der Zeitmessung.

Problematisch ist vor allem der Aspekt der Validität.

Problematisch ist vor allem der Aspekt der Validität. Während die Objektivität und die Reliabilität *formale* Kriterien sind, die etwas über die Genauigkeit und Verlässlichkeit eines Tests sagen (z.B. darüber, wie gross der Messfehler ist), ist das entscheidende Kriterium eines Tests letztlich *inhaltlicher* Art und betrifft die Frage, ob der Test überhaupt misst, was er zu messen vorgibt. Während bei einer physikalischen Messung die Frage der Validität des Messinstruments *vor* dessen Entwicklung auf der Basis von theoretischen und empirischen Kenntnissen des Messgegenstandes beantwortet werden kann, lassen sich Tests erst *im Nachhinein* validieren. Sie werden mit einem Aussenkriterium korreliert, dessen Validität aber genauso fraglich sein kann wie die Validität des Tests selber. Nur wenn wir wüssten, wie der psychologische Gegenstand beschaffen ist, liesse sich die Frage, wie er gemessen werden soll, eindeutig beantworten. Das gilt für die Erhebung psychologischer Daten mittels Tests generell, also auch für die Messung von Kompetenzen. Zwar ist uns auch die Zeit sinnlich nicht direkt

gegeben, aber im Falle der Zeit haben wir verlässliche physikalische Theorien, die uns erlauben, die Uhrzeit als fundamentalen Messvorgang zu begründen. Was die PISA-Tests tatsächlich messen, ist daher alles andere als klar.

Geringer Nutzen für die pädagogische Praxis

Offensichtlich bilden die PISA-Studien nicht einfach ab, was an unseren Schulen an Bildung vermittelt wird. Was die Studien messen, ist vom Messvorgang nicht unabhängig, sondern geht aus der Messung überhaupt erst hervor. Die Vorzüge, die die Messung von Bildung bietet, müssen daher mit einem hohen Preis bezahlt werden. Die Vorzüge liegen in der Standardisierung des Verfahrens und in seiner Unabhängigkeit von raum-zeitlichen Beschränkungen. Die Standardisierung erhöht die Verlässlichkeit, mit der über Bildung gesprochen werden kann. Die dadurch gewonnene Objektivität ist aber lediglich Objektivität im Sinne von intersubjektiver Übereinstimmung, aber nicht im Sinne von Gegenstandsadäquatheit. Zwar vermögen wir dank der PISA-Studien verbindlicher über Bildung zu sprechen, ob aber das, *worüber* wir sprechen, dem entspricht, was wir herkömmlicherweise unter Bildung verstehen, muss bezweifelt werden.

Vorteil liegt im politischen Bereich



Wie industrielle Standards dazu dienen, um Produktionsabläufe in ihrer Effizienz zu steigern, sind Bildungsstandards einer Logik der Effizienzsteigerung von Schule und Unterricht verpflichtet. Bild: AdobeStock

Angesichts dieses Vorbehalts scheint der Nutzen der PISA-Studien nicht pädagogischer, sondern politischer Art zu sein. Er liegt in der Relativierung eines sprachlichen Zugangs zur Bildung, dessen Grenzen darin liegen, dass wir über Bildung nur vage und unpräzise sprechen können, während die Messung von Bildung nicht nur mehr Präzision mit sich bringt, sondern auch Vergleiche ermöglicht, die über lokale Grenzen hinausgehen. Das entspricht dem Credo der standardbasierten Reform des Schulsystems. Wie industrielle Standards dazu dienen, um Produktionsabläufe in ihrer Effizienz zu steigern, sind Bildungsstandards einer Logik der Effizienzsteigerung von Schule und Unterricht verpflichtet. Dafür sind genaue Kenntnisse über die Leistungsfähigkeit des Systems unabdingbar. Mit den PISA-Studien wird der Blick auf die Schule vereinheitlicht, was politisch willkommen scheint. Pädagogisch ist aber etwas anderes gefragt. Wer pädagogisch handelt, tut dies nie auf dem Niveau standardisierter Abläufe, sondern immer in einer konkreten Situation unter Beachtung situativer und individueller Besonderheiten. Die Messung von Bildung steht damit im Widerspruch zur Logik pädagogischer Praxis. Dass Lehrerinnen und Lehrer mit den Ergebnissen der PISA-Studien wenig anzufangen wissen, kann daher nicht erstaunen.