

MITA BANERJEE, GREGOR FEINDT,  
HERBERT KALTHOFF UND BENJAMIN WIHSTUTZ

## Leistung im Vergleich

### Objektivierungen trotz Differenzierungskontingenz

Für das Selbstverständnis moderner Gesellschaften ist Leistung ein zentraler Begriff, denn kaum etwas scheint für ihren Bestand wichtiger zu sein, als ›Leistung zu erbringen‹ oder ›Leistungsträger‹ zu sein. Leistung indiziert Erfolg und soziales Ansehen – ökonomisch, politisch, kulturell, sportlich. Zugleich ist Leistung ein Konzept, mit dem verschiedene Dinge bezeichnet wurden und werden, wie etwa das Produkt einer Arbeit (›Dienstleistung‹) oder Anstrengung, eine finanzielle Unterstützung (›staatliche Leistungen‹) oder eine wirtschaftliche Potenz (›sich etwas leisten können‹). Der Leistungsbegriff hat also einen Mehrfachcharakter: Er ist eine Norm, deren Erfüllung in Form von individuellen Anstrengungen und Ergebnissen erwartet wird; er dient als Legitimation für Bewertungsinstanzen, Statusunterschiede durch materielle oder symbolische Erträge zuzuschreiben; und er funktioniert als Redewendung in Bezug auf symbolische und materielle Güter (Offe 1970: 42ff.). Er setzt voraus, dass einer Aktivität, einem Prozess oder einem Menschen von beobachtenden Dritten ein Wert zugeschrieben wird und diesem Wert kurz- oder langfristig eine Funktion zukommt, und zwar Menschen in Relation setzen zu können – zu sich selbst, zu anderen Menschen, zu Aufgaben, Tätigkeiten und Dingen.

Die Autor:innen dieses Beitrags gehen von drei Überlegungen aus: Erstens ist Leistung als ein allgegenwärtiges Prinzip und Merkmal moderner Gesellschaften der zentrale Mechanismus der gesellschaftlichen Verteilung von Menschen und maßgeblich an der Stratifikation der Bevölkerung beteiligt. Als Menschenbewertung organisiert sie Differenzierungen durch Zugangschancen, Möglichkeiten der Teilhabe und der sozialen Anerkennung, aber auch der Abwertung von Individuen oder Gruppen, denen keine Leistung – und oft damit verbunden – kein Vermögen oder kein Wille zur Leistung attestiert werden. Zweitens ist die Feststellung von Leistung (etwa durch organisatorische Verfahren der Objektivierung) eine lokal und situativ bedingte Konstitution, eine Erfindung: In arbiträrer Weise konstruieren die an diesen Verfahren beteiligten Evaluierer:innen das, was sie vermeintlich vorfinden, verwenden hierzu Darstellungsmedien, Messverfahren und (Alltags-)Theorien. Demnach wird eine ›Leistung‹ nicht erbracht und anschließend bemessen und bewertet, sondern sie entsteht erst durch die Bewertung einer

Tätigkeit etc., eine Bewertung, die schriftlich oder mündlich vollzogen wird. Drittens haben die historischen Verschiebungen des Wertbegriffs (hierzu Cook 2017) dazu geführt, dass Leistung als etwas Objektives gilt, das – in vielen Urkunden, Zeugnissen etc. immerzu dokumentiert – eine soziale Strahlkraft in alle Bereiche der Gesellschaft ausübt, die (fast) alle an sie glauben lässt. In diesem Sinne ist sie ein wichtiger Motor des sozialen Lebens in westlichen modernen Gesellschaften, da sie Fiktives in Wirklichkeit verwandelt.

Die Urteile, die Bewertungen zugrunde liegen, bleiben oft nicht nur unscharfe Einschätzungen, sondern werden mit einigem Aufwand und wissenschaftlichen Methoden hervorgebracht und können sich in lebenslang gültigen Zertifikaten manifestieren. Jedoch werden die Bewertungskriterien, Messverfahren oder Maßstäbe der Leistungsbewertung beständig ausgehandelt. Die Leistungskategorie ist folglich weder unabhängig von kulturellem, sozialem und technologischem Wandel zu fassen noch sind ihre Ergebnisse stets leicht zu replizieren (Distelhorst 2014; Reh/Ricken 2018). Vielmehr kann Leistung als veränderliche Relation zwischen Bewerteten, Bewertenden und Publikum verstanden werden, bei der immer wieder ausgehandelt und vorgeführt wird, was sie ausmacht und wie mit ihrer Feststellung Menschen kategorisiert und klassifiziert werden (Meier/Peetz/Waibel 2016). Zu berücksichtigen bleibt, dass weder die Bewertenden noch die Bewerteten oder das Bewertungsobjekt in diesem Prozess mit sich identisch bleiben.

Dieser Beitrag konzentriert sich darauf, wie das Konzept der Leistung, ihre Messung und vergleichende Bewertung in wissenschaftlichen, betrieblichen, schulischen und medialen Konstellationen ab Ende des 19. Jahrhunderts erprobt und objektiviert wurde. Wir vergleichen anhand von vier Fällen Diskurse, Wissenschaft und Praxisformen: die frühen Versuche der Leistungsmessung in der entstehenden Arbeitsphysiologie, der Wettbewerb am industriellen Arbeitsplatz, die schulische Beurteilungspraxis und das Phänomen des Alterns ohne Alterung. In diesen vier Fällen finden sich unterschiedliche Akzentuierungen des Leistungsbegriffs, unsere Analyse zieht ihr Deutungsangebot aus ihrem Vergleich.

Damit nehmen wir in den Blick, wie durch performative Akte der Bewertung und Beurteilung Menschen unterschieden und sortiert werden. Ziel unseres Beitrags ist es, in einem systematischen und historischen Vergleich die Spezifik unterschiedlicher Konzepte und Praktiken von Leistung herauszuarbeiten und anhand einer Kontrastierung der Fälle Mechanismen, Konjunkturen und Kontingenzen der Leistungsdifferenzierung aufzuzeigen. Wir fragen, was bei einer Unterscheidung von Menschen anhand einer zugeschriebenen Leistung jeweils konkret bewertet, ›gemessen‹ oder evaluiert wird und in welchen Sinnschichten (Hirschauer/Nübling 2021) sich Leistung manifestiert und verstetigt.

Damit verbunden ist auch die Frage nach den sozialen Formaten einer solchen Differenzierung (wie Vergleich und Wettbewerb) und nach ihren semiotischen Formen (etwa Gradualisierung und Stratifizierung), die mit diesen Konzepten von Leistung einhergehen und diese performativ verfestigen. Mit diesem kontrastiven Vergleich zeichnen wir sowohl die Entwicklung und Konjunkturen der Leistungsdifferenzierung seit dem ausgehenden 19. Jahrhundert nach als auch je nach Kontext und Funktion stark abweichende Konzepte von Leistung.

Wir verfahren in vier Schritten: (1.) Wir beginnen mit der Erfindung und Konstruktion von Leistung in der deutschen Arbeitsphysiologie um 1900 durch Messverfahren und Performanz. Diese bilden entscheidende Voraussetzungen für die moderne Vorstellung des menschlichen ›Motors‹ als Antrieb einer gesamtgesellschaftlichen Leistungsfähigkeit, aber auch für Praktiken der Humandifferenzierung und Bewertung nach Nützlichkeit, die exkludierend und menschenverachtend sein können. (2.) Dann untersuchen wir am Beispiel der tschechoslowakischen Schuhfabrik Baťa einen spezifischen Fall von Leistungsdifferenzierung in der industrialisierten Arbeit, der auf der arbeitswissenschaftlichen Bewertung aufbaute und dort Leistung von Führungskräften reifizierte. (3.) Anschließend zeigen wir aus einer systematischen Perspektive im Feld der zeitgenössischen Schule in Deutschland, wie die Skalierung der Bewertungskriterien entlang des Alters der zu Bewertenden relevant und das Notensystem von Lehrkräften als objektives Kriterium konstruiert wird. (4.) Und schließlich betrachten wir mit dem historisch ausgesprochen jungen Phänomen des ›Successful Aging‹ in den USA eine spezifische Konstellation von Leistung als Kompensation von Alter. Abschließend kontrastieren wir die vier Fallstudien und formulieren ein Fazit: Das meritokratische Modell der Leistungserbringung und -bewertung beobachtet und behandelt individuelle Körper, es singularisiert die individuelle Leistung, es differenziert ihre Träger mittels verschiedener Metriken und es stärkt durch soziale Diskurse den Glauben an das durch sich selbst Erreichte.

## 1. Messbarkeit und Quantifizierung: Leistungsfähigkeit in der Arbeitsphysiologie um 1900

Die »Erfindung der Leistung« als Paradigma der Moderne ist vor allem der zweiten Phase des späten 19. Jahrhunderts zuzuordnen, in der sich standardisierte Verfahren des Beobachtens, Messens und Vergleichens von Leistung etablieren (Verheyen 2018). Neben dem modernen Leistungssport, also dem regelbasierten und wiederholbaren körperlichen Wettbewerb, der im 19. Jahrhundert in Form von zahlreich gegründeten Sportvereinen sowie Wettkämpfen wie der Olympiade institutionalisiert

wurde, sind es insbesondere neue wissenschaftliche Experimente und Testverfahren, die menschliche Leistungsfähigkeit zu einer messbaren und vergleichbaren Größe machen. Im Folgenden soll anhand eines dieser Verfahren exemplarisch dargelegt werden, wie sich um 1900 ein neuer, von der Physik und Arbeitsphysiologie beeinflusster Leistungsbegriff etabliert, der auf Performanz, Quantifizierbarkeit und Vergleichbarkeit beruht und nachhaltige Konsequenzen für Praktiken der Humandifferenzierung in so unterschiedlichen gesellschaftlichen Feldern wie Bildung, Industrie und Freizeitkultur hat.

Eine frühe Herausforderung für die mit Leistung befassten Wissenschaften war die Frage, welche Faktoren die Leistungsfähigkeit eines Menschen beeinflussen. Der deutsche Psychiater und Arbeitsphysiologe Emil Kraepelin interessierte sich dafür, unter welchen Umständen geistige Leistungsfähigkeit abnimmt und führte diesbezüglich einige Experimente durch. In seinem 1894 veröffentlichten Aufsatz »Über geistige Arbeit« (Kraepelin 1894), in dem der Begriff der Leistung mehr als fünfzigmal verwendet wird, berichtet er über eine solche Versuchsreihe, für die er Probanden mehrere Stunden lang ohne Unterbrechung rechnen ließ, um zentrale Faktoren der Leistungsfähigkeit und »Ermüdung« des Menschen zu eruieren (ibid. 36). In eigens für den Versuch gedruckten Heften mussten die Probanden hunderte von untereinanderstehenden Zahlen addieren, wobei alle fünf Minuten ein Glockensignal ertönte, das die Versuchsperson anwies, einen Strich hinter der zuletzt addierten Zahl zu setzen, um nach Ablauf des Versuches feststellen zu können, wie viele Zahlen in je fünf Minuten von den einzelnen Personen addiert worden waren. Die Ergebnisse zeigten, dass die Rechengeschwindigkeit der Probanden (wenig überraschend) mit Dauer des Experiments abnahm – aus Kraepelins Sicht ein Beweis für die Ermüdung des Menschen bei geistiger Arbeit. Darüber hinaus stellte er fest, dass die Rechenleistung von weiteren Faktoren wie Ablenkbarkeit und Übungsfähigkeit maßgeblich beeinflusst wurde – je häufiger die Probanden an der Versuchsreihe teilnahmen, desto besser und schneller rechneten sie.

Kraepelins Erkenntnisse wiesen auf dem Forschungsgebiet der Arbeitsleistung nach, dass Leistungsfähigkeit, Übung und Ermüdung nicht nur im Bereich körperlicher, sondern ebenso bei geistiger Arbeit messbare Faktoren sind. Kraepelin interessierte sich für diese Faktoren vor allem in Bezug auf Möglichkeiten der Leistungssteigerung und -kompensation: »Der ursprünglich Langsame kann daher durch häufige Wiederholung der Arbeit allmählich den zunächst schneller Rechnenden einholen und überflügeln« (ibid.: 35). Allerdings sei jene individuelle Leistungssteigerung durch Übung endlich: »Mit jeder weiteren Wiederholung der Arbeit wird der durch sie erzielte Zuwachs an Rechengeschwindigkeit geringer. [...] Wir werden daher ziemlich bald bei jedem Menschen an eine

Grenze gelangen, jenseits derer eine Steigerung der Arbeitsgeschwindigkeit durch die Übung nicht mehr gesteigert werden kann« (ibid.). Leistung wird hier als messbare Kapazität gefasst, die sich erst über einen längeren Beobachtungszeitraum und entlang einer mehrfach wiederholten Tätigkeit zuverlässig bestimmen lässt. Das Risiko der Leistungssteigerung sah Kraepelin in der Ermüdung, die ihm, wie vielen seiner Zeitgenossen, als Gefahr am Horizont der modernen Industriegesellschaft erschien. Ermüdung wurde entsprechend nicht allein als individuelle Erschöpfung von Arbeit angesehen, sondern auch als gesamtgesellschaftliche Gefahr. »Gewinnt aber einmal die Ermüdung die Oberhand, so erfolgt das Sinken der Leistungsfähigkeit rasch und unaufhaltsam« (ibid.: 36).

Anders als in den Diskursen des 18. und frühen 19. Jahrhunderts, die die *Lebensleistung* als Errungenschaft und Erfolg in den Vordergrund stellten, wurde *Leistungsfähigkeit* von Kraepelin als begrenzte, energiebasierte, quantifizierbare Größe definiert, deren optimale Ausschöpfung von anderen Faktoren wie Nahrungsaufnahme, Ruhepausen, Übung, Arbeitszeiten und Konzentration abhängt. Menschliche Leistungsfähigkeit wurde dabei – typisch für die Zeit um 1900 – analog zur Leistung einer Maschine als »menschlicher Motor« (Rabinbach 1990) gedacht. So beginnt auch Kraepelins Aufsatz mit einem solchen Vergleich:

»Wenn heute ein Schiff seine Probefahrt macht oder der Plan einer elektrischen Beleuchtungsanlage entworfen wird, so pflegen wir mit der Befriedigung des Kulturmenschen in den Zeitungen zu lesen, wie viel indicierte Pferdekkräfte der neue Panzer zu entwickeln vermag, oder wie gross die Zahl der Bogen und Glühlampen von bestimmter Lichtstärke sein wird, [...] Bei solider Arbeit hält die Maschine genau, was ihr Erbauer versprochen hat, und er ist sogar an der Hand gewisser Erfahrungen im Stande, zu sagen, wann eine Erneuerung einzelner Teile nötig sein wird und wie hoch sich der Verbrauch an Betriebsmitteln beläuft.« (Kraepelin 1894: 31)

Bedauerlicherweise sei die Arbeitsleistung des Menschen in Bezug auf ihre Messbarkeit und Voraussehbarkeit bisher weitaus weniger erforscht:

»Wir wissen zwar ziemlich genau, welche Menge von Nahrungsstoffen dieser oder jener Organismus verbraucht, aber wir sind nur wenig im Klaren darüber, wie viel derselbe *leistet* und namentlich, wie viel er zu *leisten* im Stande ist.« (ibid.: 31, Herv. die Verf.)

Ziel von Kraepelins Versuchsreihen war entsprechend, Messverfahren menschlicher Leistungsfähigkeit zu entwickeln, die wie beim Motor oder einer Lichtanlage genau voraussagen können, wie viel ein Mensch »zu leisten« vermag, wie sehr diese Leistung gesteigert werden kann und wann der Zustand der Ermüdung einsetzt. Leistungsfähigkeit und Ermüdbarkeit sollten dabei als Orientierungsgröße nicht nur der

industriellen Arbeit zugutekommen, sondern auch mit Blick auf angemessene Lerntempi und Lernmethoden in der Schulbildung veranschlagt werden – auch aus diesem Grund maß Kraepelin die Geschwindigkeit des Rechnens. Das bisherige Schulsystem, so kritisierte er, versäume es, die Ermüdbarkeit der Schulkinder in Betracht zu ziehen und die Lehrmethoden entsprechend anzupassen. Es sei eine »unbestreitbare Tatsache, dass die Anforderungen, welche die Schule an die geistige Leistungsfähigkeit ihrer Zöglinge stellt, sehr weit über das Zulässige hinausgehen« (ibid.: 40).

Kraepelins Rechenversuche können mit anderen Experimenten und Methoden der Leistungsmessung verglichen werden, die ebenfalls von einem neuen ökonomisierten Leistungsverständnis um 1900 zeugen: Angelo Mosso entwickelt 1888 den Ergographen, ein Gewichtzug-Gerät, das die Arbeit aufzeichnet, die der Mensch mit einem Finger verrichtet, wobei Mosso insbesondere den messbare Zusammenhang von körperlicher und gedanklicher Anstrengung untersucht (Verheyen 2018: 137); 1894 veröffentlicht der Schweizer Biologe Robert Keller die Ergebnisse pädagogisch-psychometrischer Studien, die darlegen, wie sich die Leistungsfähigkeit von Schülern und ihre »Ermüdungskurven« bei Tätigkeiten wie dem Lesen oder Rechnen verändern und optimieren lassen; 1905 wird der Intelligenztest von Binet und Simon erfunden und 1911 veröffentlicht Frederick W. Taylor seine zeitbasierten Messverfahren von Produktionsleistung und Arbeitsschritten als Prinzipien des *Scientific Management*, die in den Folgejahren als Grundlage von Rationalisierungsmaßnahmen in zahlreichen US-amerikanischen und europäischen Industriebetrieben herangezogen werden. Gemeinsam ist diesen Forschungen und Publikationen, dass sie auf die Messbarkeit von Leistung als quantifizierbare, zeitlich strukturierte und ökonomisierte Größe und Kapazität zielten und sie zu objektivieren versuchen.

Während Kraepelin mit dem Glockensignal die zeitliche Struktur des Experiments als Rhythmus vorgab und Keller sogar die mittlere Zeit des Lesens pro Silbe bei seinen Probanden ausrechnete (Keller 1894: 442), maß Taylor die Optimierung der Arbeitsleistung mit der Stoppuhr (Taylor 1911: 118). Als Grundvoraussetzung für diese quantitative Objektivierung von Leistung kann daher nicht zuletzt die Standardisierung messbarer Zeit im 19. Jahrhundert betrachtet werden. Diese moderne Vorstellung einer linearen und gleichmäßig fortschreitenden Zeit setzte sich erst im 19. Jahrhundert durch die Verwendung mechanischer Uhren, aber auch durch technische Erfindungen wie die Eisenbahn und den Telegraphen durch (Oggle 2015: 75-98; Conrad 2017: 828).

Neben der zeitbasierten Messung und Quantifizierung von Leistung rückte zunehmend auch ihr performativer Charakter in den Vordergrund. Es ging in den Experimenten um Praktiken, deren Leistung weder

in der Fertigstellung eines Produkts oder Artefakts noch in der Lebensleistung im Sinne eines Verdiensts liegt. Gemessen wurden vielmehr eine Aktivität und Produktivität, die im Moment der Ausführung *hic et nunc* als Leistung beobachtbar und messbar wird, eine Performanz der Leistung. Kraepelins Leistungsbegriff kann damit als Wegbereiter eines performativen Verständnisses von Leistung gesehen werden. Denn wenn allein die Leistung *in situ* und *in performance*, also im Moment beobachtbarer Aus- oder Vorführung zählt, so kann leichter von anderen Kategorien der Humandifferenzierung mit längerer Laufzeit wie sozialer Herkunft, Ethnizität oder Geschlecht abstrahiert werden.

In diesem Sinne muss die um 1900 durch Leistungsexperimente eingeführte Vorstellung einer messbaren und quantitativ vergleichbaren Leistung auch von einem starren Verständnis der Lebensleistung (*merit*) abgegrenzt und unterschieden werden. Und es nimmt nicht Wunder, dass zur selben Zeit große Leistungsschauen wie die Olympischen Spiele oder Kraftmesser auf Jahrmärkten wie *Hau den Lukas* eine sichtbare Konjunktur erfahren (Wihstutz 2021). Die quantifizierbare, beobachtbare und vergleichbare Leistung wird in der Vorführung und Ausführung vor einer beobachtenden und urteilenden Instanz – sei es der Schiedsrichter oder das Auge eines Arbeitsphysiologen – zum Paradigma einer neuen Dimension der Humandifferenzierung. Erst vor dem Hintergrund der allgemeinen Messbarkeit, Beobachtbarkeit und Quantifizierung in so unterschiedlichen Feldern wie Leistungssport, Populärkultur und Wissenschaft wird das Wettbewerbsversprechen, dass jeder Mensch unabhängig von seiner Herkunft, Hautfarbe oder seinem Geschlecht, etwas leisten kann, glaubhaft und einlösbar. Die Bewertung von und nach Leistung erscheint so aufgrund ihrer Performanz und Quantifizierbarkeit in Kombination mit ihrer Entkoppelung von sozialen Zugehörigkeiten als gerecht.

Allerdings unterscheidet sich Kraepelins Vorstellung der Leistungsmessung gerade in diesem Punkt von einem Leistungsverständnis, das die individuelle und kompetitive Unterscheidung von Menschen nach Leistung in den Vordergrund stellt. Denn es ging ihm bei seinen Rechenversuchen weder in erster Linie um eine Differenzierung und Hierarchisierung von Menschen, noch zielten seine Experimente auf eine Typisierung von Leistungsklassen. Zwar interessierten ihn durchaus »die Unterschiede in der Ermüdbarkeit der einzelnen Personen« (Kraepelin 1894: 36), indes zielte Kraepelins Forschung wie die von Keller oder Taylor, eher auf eine kollektive als eine individuelle Steigerung der »Leistungsfähigkeit« (ibid.: 40), insbesondere auf eine Vermeidung von Ermüdung. Entsprechend sprach er sich, wiederum ähnlich wie Taylor, für mehr Ruhepausen und für gesunde Ernährung aus. Auch interessierte ihn Leistung, anders als später häufig beim IQ-Test (Hegarty 2013), nicht im Vergleich von Geschlechtern oder Rassen. Zumindest in Kraepelins Aufsatz »Über geistige Arbeit« scheint die messbare Leistungsfähigkeit daher

noch nicht als Form der Humandifferenzierung etabliert. Vielmehr wird sie als eine Art Motor der Gesellschaft (bes. auch der Bildung) gedacht, für den der richtige Treibstoff, die optimale Auslastung und die Verhinderung von Verschleiß über die Messverfahren erst noch genauer entwickelt werden müssen.

Die Messbarkeit und Quantifizierbarkeit von Leistung impliziert somit unterschiedliche Verfahren der Objektivierung und Bewertung, die weitreichende Folgen für die Unterscheidung von Menschen haben können. Während Kraepelin, Keller oder Taylor sich noch einem sozialdemokratischen Fortschrittsdenken verpflichtet sahen, das die Leistung des menschlichen Motors unter Rücksichtnahme auf Arbeiterrechte und Erholung behutsam zu steigern versuchte, wurden nur wenige Jahrzehnte später Menschen in Deutschland nicht nur nach Rasse und Gesundheit, sondern auch rigoros nach Leistungs(un)fähigkeit kategorisiert und zu »lebensunwertem Leben« und »Ballastexistenzen« erklärt.<sup>1</sup> Der arbeitsphysiologisch begründeten Messung, Quantifizierung und vergleichenden Bewertung von Leistungsfähigkeit eignet somit sowohl ein Gleichheitsversprechen (allein die Leistung zählt und zwar im Moment der Messung) als auch die Möglichkeit einer Verdauerung von Humankategorien (leistungsstark vs leistungsschwach, hochbegabt vs lernbehindert etc.), die sich über wiederholte, objektivierte Leistungsmessungen (beispielsweise IQ-Tests) legitimiert.

## 2. Reifizierungen:

### Messen und Bewerten als Mittel der Personalpolitik

Leistung lässt sich auch in der modernen industriellen Arbeitswelt im Spannungsverhältnis zwischen Arbeit und Zeit beobachten. Dabei wird diese Leistung im Arbeitskontext nicht auf gesamte Gesellschaften bezogen, sondern vielmehr an einzelnen arbeitenden Menschen oder kleineren Gruppen von Arbeitenden festgemacht und diesen als feststellbare Eigenschaft zugeschrieben. Im Ersten Weltkrieg begannen große

- 1 Bemerkenswerterweise wurden die NS-Krankenmorde (die sogenannte »Euthanasie«) nicht allein eugenisch, sondern ebenso volkswirtschaftlich mit einem Leistungsdiskurs begründet. Dies lässt sich in der bereits 1920 veröffentlichten für den Nationalsozialismus einflussreichen Schrift »Die Freigabe der Vernichtung lebensunwerten Lebens« von Alfred Hoche und Karl Binding nachlesen, in der erstmals von behinderten und psychisch kranken Menschen als »Ballastexistenzen« die Rede ist. Die Schrift plädiert für die Morde an unheilbar Kranken, um ein »Freimachen jeder verfügbaren Leistungsfähigkeit« (Hoche/Binding 1920: 50) zu gewährleisten und die Vergeudung von Arbeitskraft in der Pflege zu vermeiden.

US-amerikanische Unternehmen wie Ford, aber auch etwa Siemens in Deutschland, damit ihre Personalpolitik zu systematisieren und bürokratisch zu organisieren (Patzel-Mattern 2010) und trugen mit arbeitswissenschaftlichen Experten grundlegend zur »Verwissenschaftlichung des Sozialen« bei (Raphael 1996). Dabei diente Leistung als vermeintlich objektives Kriterium für die Entlohnung von Beschäftigten oder Personalentscheidungen und wurde bei Führungskräften vielfach als gegeben und unverhandelbar betrachtet (Kaufman 2010: 87–102; Luks 2010: 146–155). Im Unterschied zur als Akkord quantifizierbaren Arbeit der Werkarbeiter in der Fabrikhalle erschien die Leistung von Verwaltungsmitarbeitern und Führungspersonal als Eigenschaft der Menschen; ein dem Menschen inhärentes Verständnis von Leistung wirkt also weiter.

Ein prägnantes Beispiel dieser Entwicklung war das tschechoslowakische Schuhunternehmen Baťa in Zlín, das als erster Hersteller in Europa die rationalisierte und vollständige maschinisierte Produktion von Schuhen einführte und in den 1920er und 1930er Jahren global expandierte. Orientierte sich das Zlíner Unternehmen anfangs noch an US-amerikanischen Vorbildern, überholte es diese in den 1930er Jahren bei der bürokratisierten Personalverwaltung und entwickelte eine eigene Form des Personalmanagements, dass der Entwicklung in den USA und West-Europa deutlich vorausgriff. Baťa übertrug die Modelle und Organisationsformen der Rationalisierung auf die bürokratische Behandlung von Beschäftigten und entwickelte ein aufwändiges *social engineering* (Etzemüller 2009) der eigenen Belegschaft mit dem Ziel, leistungsfähige, in der Regel männlich gedachte, Beschäftigte zu identifizieren und bestmöglich für Produktion und Vertrieb einzusetzen (Feindt 2017; Doleshal 2021).

Die Messung und Bewertung von Arbeitskräften folgte bei Baťa einer Eigenlogik, die auch in der Arbeitswissenschaft oder im Sport zu beobachten ist. Demnach entsteht Leistung durch den zu bewertenden Menschen selbst, beruht auf einem individuellen Leistungspotential, das abhängig von verschiedenen Faktoren abgerufen werden kann. Entlang dieses Verständnisses nutzte Baťas Personalverwaltung zunächst neuartige quantitative Verfahren, um individuelle Beschäftigte und ihre Arbeitsergebnisse zu vergleichen. Nach dem Ersten Weltkrieg führte das Unternehmen Intelligenztests ein, die mit mathematischen Aufgaben die Fähigkeiten und das intellektuelle Potential von Menschen vor allem für eine bestimmte Tätigkeit feststellen sollten, und machte ein wenigstens zufriedenstellendes Ergebnis zur Bedingung für eine Einstellung, auch bei handwerklichen Tätigkeiten. Wichtiger aber noch wurde der Lohnvergleich, für den das Unternehmen nicht erst mit wissenschaftlichen Methoden Zahlen erzeugen musste, sondern sein ohnehin verfügbares Accountingwissen nutzte. Baťa zahlte in der Produktion individuelle bzw. gruppenbezogene Akkordlöhne und beteiligte Meister:innen

und Abteilungsleiter:innen am Gewinn ihrer Werkhalle. Bei der wöchentlichen Lohnauszahlung erläuterten Führungskräfte die Berechnung dieser Löhne und ihre Unterschiede und demonstrierten so den direkten Zusammenhang von Arbeitsergebnis und Differenzierung vor der Belegschaft der jeweiligen Werkhalle als berechenbares, geordnetes und transparentes Verfahren (Steinführer 2002: 49f.).

Ab 1926 inszenierte das Unternehmen seine Löhne auch in der städtischen Öffentlichkeit und druckte in den Betriebszeitungen punktuell die höchsten Löhne und Gewinnbeteiligungen der Woche, zunächst für die erfolgreichsten Abteilungen und ab 1931 regelmäßig auch für Einzelpersonen, die mit Namen, Bild oder einem kurzen Porträt vorgestellt wurden. Das Unternehmen stellte diejenigen Männer und Frauen heraus, die – so die Betriebszeitungen – »bei ihrer Arbeit führen« und deren »Redlichkeit und Qualität bei der Arbeit sich selbstverständlich in der Höhe ihrer Einnahmen widerspiegeln«. Es proklamierte den Grundsatz, »dass die Besten die höchsten Löhne erhalten« (Zlín, 12.08.1932). Der simplifizierende Vergleich von Menschen anhand ihrer Lohnsumme, der produzierten Stückzahlen bzw. der möglichst geringen Menge an Ausschuss machte Leistung ähnlich wie in der oben diskutierten Arbeitsphysiologie zu einer quasi-physikalischen, messbaren Größe. Die Unterscheidung nach Leistung erschien faktenbasiert, effizient und gerecht und beanspruchte Eindeutigkeit (Verheyen 2018: 54–57). Auch wenn die Bestimmung der Zahlen »eine Vielzahl von Entscheidungen erfordert[e]«, wurden diese von Klassifizierern wie auch Klassifizierten »für die Sache selbst gehalten« (Heintz 2010: 170).

Ein solcher numerischer Vergleich stieß jedoch bei Führungspositionen an seine Grenzen. Zwar identifizierte Baťa anhand von überdurchschnittlichen Löhnen besonders produktive Beschäftigte, konnte diese Logik aber nicht fortsetzen, wenn diese durch Beförderung von der Produktion in die Verwaltung wechselten, d.h. nicht mehr einer quantifizierbaren Tätigkeit nachgingen und die Unterscheidung von Arbeitern und Angestellten überschritten hatten (Weischer 2017: 93). 1934 führte das Unternehmen dazu Jahresbewertungen ein, für die die Personalabteilung u.a. die Kategorien ›Arbeit‹, ›Betragen‹, ›Charakter‹, ›Führungsqualitäten‹ und ›Verhältnis zum Betrieb‹ mit Noten bewertete. Auch wenn messbare Leistung in der tschechoslowakischen Arbeitswissenschaft intensiv diskutiert wurde (Janda 1932–34), erschien sie hier nicht als Bewertungskriterium.

Leistung im Sinne einer durch Bewertung hervorgebrachten Distinktion wurde durch die *Summe* der Bewertung angezeigt. Dabei formulierten die (zumeist männlichen) Vorarbeiter, Meister und Vorgesetzte als Machttechniker des Unternehmens die Noten in jeweils eigenen Attributen, die einer vierstufigen Rangliste folgten. Abschließend stellte die Bewertung fest, ob Beschäftigte sich »im Aufstieg« befänden, stagnierten

oder in ihrer Leistung »im Abstieg waren« und im Ausnahmefall sogar unmittelbar zu entlassen seien. Hier evaluierten die Vorgesetzten und Personalreferenten auch extrafunktionelle Aspekte der Arbeit, also solche die mit der Effizienz der Produktion nur bedingt verbunden waren, und im weitesten Sinne solche, die das Sozialverhalten der Beschäftigten und deren Erfüllung einer Arbeitsrolle innerhalb der industriellen Organisation erfassten (Offe 1970: 29). Innerhalb der Sinnwelt des meritokratischen Schuhunternehmens waren es solche extrafunktionellen Faktoren, besonders die Loyalität zum Unternehmen und die Konformität der Lebensführung, die zur Voraussetzung für beruflichen Aufstieg werden sollte.

In einem weiteren Schritt führte das Unternehmen Mitte der 1930er Jahre für erfahrene Führungskräfte Persönlichkeitstests ein. Anders als die üblichen Intelligenztests erhoben diese von arbeitswissenschaftlichen Instituten durchgeführten Evaluierungen keine quantifizierten Daten. Sie erstellten ein Profil der jeweiligen Person und positionierten diese zu ihrer Tätigkeit und anderen Beschäftigten. Auch betonte der im Test enthaltene Intelligenztest nicht den erreichten Zahlenwert, sondern die spezifische Eignung für die ausgeübte Tätigkeit. Unter der Rubrik Leistungsfähigkeit fragten die Experten Anpassungsfähigkeit, Ideenreichtum, Initiative, Organisationsfähigkeit und Ausdauer ab und verbanden dies mit detaillierten Reflexionen über Stärken, Schwächen und die gesundheitliche Eignung der Person. Eine solche »Klassifikation der Klassifizierer« (Bourdieu 1984) vertraute auf externe Experten und verwendete ein gänzlich anderes Konzept von Produktivität als bei der Bewertung einer Arbeit am Fließband. Beschäftigte in Leitungspositionen und Machttechniker wurden gerade nicht über eine situativ erbrachte Leistung oder ein variables Arbeitsquantum definiert, sondern ausschließlich über qualitative Kriterien wie ihren Charakter, also über vermeintlich beständige Eigenschaften und die daraus resultierende dauerhafte Leistungsfähigkeit. Dieser Wechsel von einer numerischen Objektivierung hin zu einer sozialen Objektivierung verortete Beschäftigte im ideologischen Horizont des Unternehmens, wenn etwa männliche Erzieher in der betriebseigenen Berufsschule, die selbst weder in Produktion noch Vertrieb eingesetzt wurden, für ihren Unternehmergeist und ihr Streben nach mehr wirtschaftlicher Tätigkeit belobigt wurden.

Eine solche Beschreibung von »Persönlichkeit« reifizierte die Arbeitsproduktivität und den Wert eines Beschäftigten für das Unternehmen. Hatte bereits das jährliche Urteil einzelne Bewertungen in Form von Personenkategorien fixiert, etwa Beschäftigte mit der höchsten Bewertung in der Kategorie Führungsqualitäten als »Organisator« herausgestellt, verstärkte sich diese charakterliche Zuschreibung und Reifizierung in den Persönlichkeitstest. Wer in den Bewertungen sehr gut abschnitt, im Aufstieg begriffen war und wem im Persönlichkeitstest herausragende

Fähigkeit und nur wenige Schwächen attestiert wurden, verrichtete nicht bloß sehr gute Arbeit, er – in der Regel wurde diese Einschätzung nur für Männer vergeben – war ein sehr guter Beschäftigter. Die Bewertung ging auf den Bewerteten über und wurde zu einem Teil von ihm (Parsons 1937/1991: 41f.).

In diesen Bewertungen von Beschäftigten wird die soziale Strahlkraft von Leistung im Industriebetrieb deutlich: Auch wenn vielfältige, zum Teil widersprüchliche Objektivierungen das Konzept Leistung konkretisierten, imaginierten die Inszenierungen der Lohnauszahlung, Auszeichnungen, Beförderungen und anderen Momenten der Abstandsvergrößerung zwischen Beschäftigten einen übergreifenden Leistungsbegriff. Diese lokal und situativ bedingte Inszenierung betonte die Rationalität und Modernität industrieller Arbeit und setzte damit Leistung als Unterscheidungskriterium in einen immanenten Gegensatz zur von traditionellen Eliten, Handwerk und Landwirtschaft geprägten Gesellschaft der Stadt Zlín und des sie umgebenden östlichen Mährens. Baťa versprach – ähnlich wie andere Unternehmen der Zeit – seinen Beschäftigten die Möglichkeit des sozialen Aufstiegs und der Teilhabe an einem urbanen, komfortablen und modernen Leben. Das Beispiel verdeutlicht, wie die Vorstellung einer rational und objektiv bewerteten Leistung dem Individuum Zukunft und Wohlstand verspricht und zu einem wirkmächtigen Glaubenssatz des sozialen Zusammenlebens wird.

Doch wie ging eine solche meritokratische Differenzierung mit Autoritäten um, deren Leistung gar nicht infrage gestellt werden durfte? Baťa erhob den kontinuierlichen Aufstieg von der Fabrikarbeit zur Führungsposition zum Ideal der Arbeitskarriere im Unternehmen. Wer jedoch einmal aufgestiegen war, stieß an eine gläserne Decke. War nämlich die Mehrzahl der Direktoren des Unternehmens bereits vor dem Ersten Weltkrieg oder in den frühen 1920er Jahren zu Baťa gekommen und hatte seitdem ein enges familiäres Netzwerk um den Firmengründer geknüpft, konnte in den 1930er Jahren faktisch niemand mehr in diesen Kreis aufsteigen. Für diese Direktoren wurden *pro forma* Jahresbewertungen erstellt, in denen ihre positiven Charaktereigenschaften hervorgehoben wurde, nicht aber Persönlichkeitstest – denn dies hätte die Einbeziehung von externen Experten erfordert. So beschrieb etwa die Akte des Direktors Dominik Čipera, zunächst Prokurist und ab 1932 faktisch zweiter Mann im Unternehmen, seinen selbstlosen Einsatz beim Löschen eines Brandes in der unternehmenseigenen Chemiefabrik. In einer nur zu symbolischen Zwecken angelegten Personalakte wurde über ihn vermerkt, dass er sich »durch eine außerordentliche Arbeitsleistung und durch herausragende charakterliche und staatsbürgerliche Tugenden« auszeichnete, ohne dafür weitere Belege anzuführen. (Čipera o.J.: 9) Eine solche affirmative Feststellung hob Čipera als idealen Beschäftigten heraus, ohne ihn einer quantifizierenden oder charakterlichen Bewertung

zu unterziehen, wie sie für Arbeiterinnen, technische Angestellte oder Verwaltungsangestellte erwartbar gewesen wäre und die variable Leistungen hätte feststellen können. Arbeitsleistung und Charakter wurden in Ciperas Fall lediglich postuliert, zusätzlich stach die persönliche Nähe zum Firmengründer als eigenes Kriterium heraus. Die familiär verbundene Führungsgruppe des Unternehmens akkumulierte so neben wettbewerblichen Leistungsfeststellungen Verdienste und Legitimationsmarker, die sie von anderen unterschied und ihre Sonderbehandlung rechtfertigte. Die herausragenden Eigenschaften waren den Mitgliedern der Leistungsgruppe geradezu körperlich und über lange Dauer eingeschrieben und setzte sie scharf von der übrigen Belegschaft ab.

Am Beispiel Ba'ā lässt sich zeigen, dass Menschen eben nicht nur anhand von Arbeitsproduktivität unterschieden wurden, sondern dass soziale Dimensionen in die Bewertung einfließen und damit eine soziale Hierarchie der Positionen und der Arbeitsbereiche (Produktion vs. Verwaltung) offensichtlich machte. Je höher die Position in der Hierarchie, umso mehr wurde belobigt und eine unspezifische Leistung gewürdigt, deren Ergebnis schon rekursiv in der Position angelegt war. In der bürokratisierten Personalverwaltung wurde Leistung für Arbeiter:innen in der Produktion quantifizierend entlang des Akkordlohns evaluiert, für qualifizierte Kräfte in Form von Noten erhoben und für Führungskräfte mit wissenschaftlichen Methoden als Persönlichkeitsprofil gefasst. Dieser Übergang von einer rationalisierten, aus der Produktion abgeleiteten Bewertung zu einer Zuschreibung von positiven Eigenschaften verschob den Begriff der Leistung von einer kleinteiligen, wiederholbaren Tätigkeit hin zu einer verallgemeinerten, als Charaktereigenschaften gefassten Zuschreibung. In der Fabrikhalle war Leistung Quantität und Qualität auf Zeit; Zugehörigkeit zum Führungsgremium wurde durch den Menschen immanenten Eigenschaften dauerhaft begründet.

Mit diesen unterschiedlichen Leistungsbegriffen ging zugleich eine Stratifizierung der Leistung einher. Standen nämlich die zu bewertenden Beschäftigten in der Handlungslogik der Personalabteilung untereinander in einem direkten Wettbewerb, der persönlichen Aufstieg versprach, verflachte diese Konkurrenz unter den Machttechnikern bereits und fiel im herausgehobenen familiären Netzwerk der Direktoren gänzlich aus. Wurden also bei den unteren Rängen des Unternehmens ein neuartiger, an den oben ausgeführten wissenschaftlichen Studien orientierter Leistungsbegriff eingesetzt, galt für Führungskräfte und deren familiales Netzwerk ein älterer Leistungsbegriff. An die Stelle von Leistung als Wettbewerb trat vielmehr die Performanz bestimmter Tugenden und Eigenschaften, die festgestellt werden konnten oder Verdienste, die im Fall von Direktoren schlicht behauptet wurden. An die Stelle einer rationalisierten und ergebnisoffenen Evaluierung trat die Heraushebung und Dissimulierung der Machttechniker und Führungskräfte bis hin zu

deren Unvergleichbarmachung. Das Beispiel Ba'ta verdeutlicht damit die hohe Valenz von Leistung in einem Industrieunternehmen und zugleich die Ambiguität eines Begriffs, der sowohl zur Hervorbringung als auch Begründung hierarchische Unterscheidung dient.

### 3. Schule und die Kontingenz ihrer Leistungsfeststellung

Betrachten wir einen dritten Fall mit einem besonderen Augenmerk auf die Konstruktivität der Leistungsbewertung. Die Schule ist die Organisation, die die Funktion einer systematischen Bildung und Erziehung nachwachsender Generationen erfüllt, sie zu Gesellschaftsmitgliedern macht und durch fortlaufende Leistungsbewertung auf weitere Leistungsbewertung einstellt. Die Forschung zur schulischen Bewertung ist dabei ebenso Teil des Erziehungssystems wie die arbeitswissenschaftliche Forschung Teil von Arbeitsprozessen ist. Schauen wir uns also sowohl die schulische Bewertungspraxis wie den wissenschaftlichen Diskurs über diese Praxis an.

Die schulische Bewertungspraxis gründet auf einer Reihe von Annahmen, deren Gültigkeit sehr umstritten sind: Erstens geht man davon aus, dass gute Noten ein knappes Gut sind und Schüler daher nicht alle die gleiche gute Bewertung bekommen können, Noten also streuen müssen. Erst die (Normal-)Verteilung der Schüler über die Notenskala macht die Notengebung zu einer sinnvollen Bewertung. Zweitens geht man davon aus, dass eine Leistung der Schüler:innen vorliegt und Lehrkräfte das, was sie ›die Leistung‹ nennen, nur noch abbildend beurteilen. Diese Annahme lässt sich darauf zurückführen, dass (a) von einer Linearität des Unterrichtsgeschehens ausgegangen wird (Lehren – Lernen – Prüfen – Bewerten), und dass (b) von allen Umständen abgesehen wird, die ›Leistung‹ nicht isolierbar und individuell zurechenbar machen. ›Leistung‹ ist der Begriff, der das schulische Geschehen dekontextualisiert und eine Zurechnung auf die Aktivitäten und Potenziale der Schüler:innen organisiert. Drittens gilt ein meritokratisches Prinzip als ›Alchemie der Schule‹ (Bourdieu 1989): Sozial oder kulturell ungleiche Schüler:innen werden als fiktive Gleiche versämtlicht, um sie dann durch die Bewertung als schulisch Ungleiche markieren zu können. Dieses Prinzip folgt der Annahme, dass der Unterricht den Wissensstand der Schüler im Prinzip vereinheitlicht.

Diese hier nur skizzierten Annahmen (siehe hierzu Filer/Pollard 2000; Brookhart 2004; Rabenstein et al. 2013; Kalthoff 2017) sind seit Ende des 19. Jahrhunderts (etwa Schreiber 1899) immer wieder Gegenstand sozialwissenschaftlicher Kritiken gewesen, die u.a. die Kontingenz des schulischen Urteils belegt haben. In den letzten 20 bis 30 Jahren hat sich

die Forschung sowohl empirisch als auch theoretisch mit dem Bewertungsprozess auseinandergesetzt (z.B. Brookhart 1993; Stiggins/Conklin 1992; Zaborowski/Meier/Breidenstein 2011). Diese Studien zeigen, dass die Ergebnisse von Leistungsbewertungen und Benotungen nicht den Kriterien der Objektivität, Zuverlässigkeit und Validität entsprechen können. Stattdessen sind sie in all ihren Dimensionen kontingent (selbst in naturwissenschaftlichen Fächern und in der Mathematik) und von den historisch-kulturellen und lokalen Bedingungen jeder Schule beeinflusst. Diesen konstruktivistischen und praxistheoretischen Studien steht eine realistische Bewertungsforschung über messbare »Schulleistung« gegenüber (etwa aus dem Kontext der pädagogischen Psychologie). Ihre Arbeiten attestieren den Lehrerurteilen u.a. eine gute diagnostische Kompetenz (Südkamp et al. 2008), erkennen – trotz starker Streuung der Urteile (etwa bei Aufsätzen) – dennoch eine vergleichbare Bewertung von Erst- und Zweitbewerter:innen (Birkel/Birkel 2002) und gehen davon aus, dass sich das Urteil der Lehrkräfte gut für den klasseninternen Vergleich der Schüler:innen eignet (Hochweber 2010). Diese Studien plädieren dafür, die Urteile der Lehrkräfte genauer und zuverlässiger zu machen. Zu diesen Bemühungen zählen u.a. Bestrebungen, die Diagnostik auf ein höheres Messniveau zu bringen, um dadurch subjektiv gefärbte Urteile möglichst zurückzudrängen sowie Wege der intersubjektiven Validierung zu erproben (etwa Jürgens/Sacher 2008; Shay 2004). Vorausgesetzt aber bleibt dabei die Annahme, es handele sich bei der schulischen Prüfung um ein *Messverfahren*, bei dem diejenigen, die messen, austauschbare Instrumente sind, die zu gleichen und somit gültigen Ergebnissen kommen (müssten), da sie lediglich die Leistungen der Schüler:innen abbilden.

Die neuere Forschung (z.B. Filer 2000; Brookhart 2004) betont dagegen, dass die Bewertung in die jeweilige Lehr- und Lernkultur einzuordnen und als Zusammenspiel von Lehrkräften und Schüler\*innen zu verstehen ist (Pryor/Torrance 2000). Ferner hat sie gezeigt, dass sich die Urteilspraxis – in Relation zum realistischen Szenario – kreativ, erfindungsreich, relativierend und zuweilen auch subversiv verhält. Daher lässt sich die realistische Spielart der Forschung als Teil des Feldes und als eine Arbeit an der Formatierung der Bewertungspraxis fassen: Sie beschreibt, wie sie sein sollte. Werfen wir zum Vergleich einen kurzen empirischen Blick auf das alltägliche schulische Bewerten am Beispiel des Unterrichtsgesprächs und der Korrekturarbeit von Lehrkräften. Aus einem ethnographischen Protokoll:

Eine Lehrkraft beginnt mit ihrer Klasse ein neues Thema: die Konstruktion von Dreiecken. Sabine, die bislang eher zu den ruhigen Schülerinnen gehörte, beteiligt sich lebhafter am Unterrichtsgeschehen, zeigt häufiger auf oder meldet sich sogar dann, wenn die anderen die Antwort nicht wissen. Ihr scheint der neue Stoff regelrecht Spaß zu machen. Nach einer

Woche kommentiert der Lehrer: »Ich will nicht immer die gleiche« oder »Sabine hatten wir jetzt schon ein paar Mal«. In einer Stunde kommt es zu folgender Sequenz: Bei der Berechnung des Flächeninhaltes von Parallelogrammen übt der Lehrer mit den Schülern Formelumwandlungen. Er fragt: »Wie kommt man auf die Höhe  $h_a$ ?«. Nur Sabine zeigt auf, wird ausgewählt und sagt: »A durch a ((A:a))«. »Richtig«, antwortet der Lehrer, »Wie kommst du drauf?« Sabine leitet die Formel über die Konstruktionszeichnung her, die an der Tafel steht. Aber noch während ihrer ausführlichen Antwort wird sie vom Lehrer unterbrochen: »Du sollst es nicht geometrisch herleiten, sondern rechnerisch«. Mit dieser deutlichen Fehlerkennzeichnung beendet der Lehrer ihr Rederecht; er argumentiert nicht, erklärt auch nicht mehr seine Frage, sondern bricht ab. Für den weiteren Klärungsprozess der Ausgangsfrage (»Wie kommt man auf die Höhe«?) nimmt er einen Schüler dran, der in der Klasse als sein »Lieblingsschüler« gilt. Mit ihm organisiert er Frage um Frage, Stichwort um Stichwort die rechnerische Herleitung der Formel.

Diese Szene bietet sicherlich verschiedene Interpretationsmöglichkeiten; von Interesse ist hier die Frage, wie es der Lehrer schafft, dass eine »Dreierschülerin« trotz vieler richtiger Antworten eine Schülerin bleibt, die er weiterhin mit Drei bewerten kann. Das vorliegende Interaktionsmuster zeigt, dass sie sich aktiv an der Erschließung des neuen Themengebietes beteiligte, und zwar mit ausführlicheren Einlassungen. In der protokollierten Szene nimmt der Lehrer der Schülerin aber das Rederecht ab und beteiligt einen anderen Schüler, der dann alle Zeit der Welt bekommt. Durch die Markierung seiner Antwort relativiert der Lehrer das Wissen der Schülerin und kann sie in ihrem mündlichen Notenbereich belassen. Es scheint hier so, als würde eine Antwort, die explizit als inadäquat markiert wird, umso rigider geahndet werden, wenn sie nicht den Vorstellungen der Lehrkraft über das erwartbare Niveau der Schüler:innen entspricht. Daraus folgt: In mündlichen ad-hoc-Kategorisierungen von Schüler:innen aktualisieren sich überdauernde Wahrnehmungs- und Handlungsschemata der Lehrkräfte – Schemata, die auf Erwartungen, vergangene Erfahrungen und auf bewährte Muster bezogen sind. Diese Form der Verdauerung überträgt sich in die Schrift der Bewertung und des Zeugnisses.

Lehrkräfte beobachten aber auch sich selbst. So meinten zwei Lehrer:innen zu ihrer Bewertung von schriftlichen Klassenarbeiten:

»Du guckst dann immer wieder. Da hat er vielleicht noch was richtig. Also was er schreibt ist völlig falsch oder ziemlich falsch, aber du akzeptierst es doch.«

»Ja gut, da ist der Zusammenhang. Ich weiß, sie meint eigentlich, es ist aber noch die falsche Fachsprache. Da kann ich aber nochmal 'n halben Punkt geben. Das kann ich machen.«

Offensichtlich können Lehrkräfte aus dem, was für sie eigentlich nicht bewertbar ist, etwas machen, indem sie einen Wert – und seien es 0,5 Punkte – zuschreiben. In diesen Fällen ihrer wiederholten interpretierenden Lektüre verschieben sie nicht nur die Differenz von ›noch bewertbar‹ und ›nicht mehr bewertbar‹, sie schützen auch sich selbst vor den Geltungsansprüchen ihrer Schüler:innen und deren Familien, indem sie in unklaren Fällen auch kaum richtige Antworten ›gut genug‹ finden. In solchen Situationen nimmt ihr Wohlwollen (aber auch ihre Missgunst), d.h. ihre Stimmung, in der sich die Relation zu ihrer Klasse spiegelt, Einfluss auf ihre Bewertung. Schließlich deutet vieles darauf hin, dass Lehrkräfte sich auch selbst in Beziehung zu ihrer Unterrichts- und Prüfungspraxis stellen. So meinte eine Lehrkraft während der Korrektur einer Geschichtsklausur (Oberstufe):

»Und zur Gewichtung der Klausur sag ich jetzt auch schon: Sollte der erste Teil wesentlich besser gelöst sein von den Schülern und der zweite ganz marode, und ich *dro:he* mit der Hälfte der Leute unter fünf zu kommen, dann erlaube ich mir nachträglich, meine Gewichtungen, meine Kriterien zu verschieben.«

Die Drohung, dass die Klausur scheitern könnte, ist eine Bedrohung des Lehrer-Selbst sowie der Annahme, Lehrkräfte kennen die Inhalte ihrer Fächer und das Wissen ihrer Schüler:innen genau. Die Option, vorsorglich Verschiebungen an Fragestellungen, an Maßstäben, an Kriterien und Ergebnissen vorzunehmen, steht für die Relevanz, die die Bewertung von Schüler:innen für die Lehrkräfte selbst besitzt. Die ›Hälfte der Schüler:innen unter fünf‹ *kann* von Lehrkräften als eine substantiell empfundene Bedrohung gelesen werden, der sie vorbeugen wollen und können. Ob sie diese vorsorgend-vorgreifende Anpassung vollziehen und damit eine Neubewertung der Bewertung oder eine Korrektur der Korrektur nachträglich umsetzen, ist u.a. von ihrem doxischen Selbstverständnis als Lehrkraft, von ihrem Verhältnis zu ihrer Klasse und zu ihrem Unterricht abhängig. Dieser Umstand, dass ein Bewertungsverfahren die Evaluierenden selbst zum Objekt der Wertung macht, steht dem Ideal einer objektiven Messung, die durch nichts gestört wird, diametral entgegen. Zwar sind in manchen Sportarten Schiedsrichterentscheidungen umstritten, aber es existiert keine Wertung der Wertenden, die ihrerseits Einfluss auf das Geschehen nimmt. Auch sind die Ermessensspielräume der Lehrkräfte – eine Art Toleranzbereich (s.o.) – etwa in der Leichtathletik undenkbar.

Diese empirischen Vignetten in Rechnung gestellt, zeichnen sich vier Neuorientierungen innerhalb der schulischen Bewertungsforschung ab. *Erstens* wird vorgeschlagen, von der Auffassung abzurücken, es läge eine Leistung der Schüler:innen vor, die von der Lehrkraft nur noch zu bewerten ist. Dagegen wird eine anerkennungstheoretische Auffassung

vertreten: Es liegt erst dann eine Leistung vor, wenn eine Arbeit, eine Anstrengung etc. von autorisierten Lehrkräften als eine solche anerkannt wird. In ihren situativen Akten der Bewertung verwandeln Lehrkräfte die Arbeit, Anstrengung etc. der Schüler:innen durch ein Zeichen (eine Note) in eine Leistung. Dieses Zeichen steht dann für den *schulischen Wert* eben dieser Arbeit oder Anstrengung und auch für ihre Autor:innen, die mit ihm identifiziert werden. Das hieße, dass die Bewertung die Leistung konstituiert (Luhmann 2002). Eine solche anerkennungstheoretische Sicht öffnet den Raum u.a. für humandifferenzierende Praktiken des Auf- oder Abwertens, des Lobens oder Demütigens, des Ignorierens oder Berücksichtigens. *Zweitens* wird vorgeschlagen, lineare Vorstellungen schulischer Phasen von Lehren-Lernen-Prüfen-Bewerten aufzugeben und das rekursiv-reflexive Verhältnis dieser Phasen in den Blick zu nehmen: etwa den vorgreifenden Rückgriff der schulischen Prüfung auf den Unterrichtsinhalt (Breidenstein 2023). *Drittens* wird vorgeschlagen, von den Gütekriterien, die für numerische Verfahren der standardisierten Sozialforschung gelten, systematisch Abstand zu nehmen: Nicht das Sein-Sollen der Objektivität des Noten-Urteils ist relevant, sondern die Frage, wie Leistung organisatorisch durch die Schule objektiviert und von Lehrkräften legitimiert wird (Verkuyten 2000). *Viertens* gerät das Tun der Lehrkräfte in den Blick. Im Akt der Fremdbewertung reflektieren Lehrkräfte somit über sich selbst und bewerten, indem sie eine Klassenarbeit von Schüler:innen bewerten, auch ihren Anteil am Zustandekommen dieser ›Leistungen‹. Gingen diese Vorschläge aus der Forschung in die schulische Bewertungspraxis ein, würde die Praxis der Selbstabsicherung und Selbstbewertung von Lehrkräften nicht länger als ein neutrales, nicht involviertes Handeln verstanden (Kalthoff 1996).

#### 4. Altern als Leistung: Successful Aging und die neue Gerontologie

Das »Successful Aging« ist, anders als die Schule, keine Institution und anders als die Arbeitswelt kein funktionaler Teil der Gesellschaft. Es handelt sich vielmehr um einen Diskurs, der sich anschickt, unsere Vorstellung vom Altern – genauer: vom Erreichen eines höheren Alters – zu verändern. Dieser Diskurs versucht seine Überzeugungskraft daraus zu erlangen, dass er sich an die Felder der Schule (›lebenslanges Lernen‹) und der Arbeitswelt (lebenslange Produktivität) anlehnt und einen Teil seiner Beschreibungen aus ihnen entlehnt. Zugleich stellt das Phänomen des Successful Aging eine Anomalie dar. Denn der dominante Altersbegriff scheint ja einen Gegensatz zu Leistung und Erfolg zu implizieren, nämlich dass Alter nicht ›gemacht‹ oder hergestellt werden kann, gilt es

doch als ein biologischer Prozess, der unaufhaltsam voranschreitet und die körperliche und geistige Leistungsfähigkeit einer Person schwinden lässt, ohne dass sie darauf Einfluss nehmen könnte.

Auch die Verbindung zwischen Altern und Leistung nimmt ihren Anfang in einem wissenschaftlichen Diskurs, der sogenannten ›Neuen Gerontologie‹. In ihrem Buch *Successful Aging* (1998) fordern der Mediziner John Rowe und der Psychologe Robert Kahn eine grundsätzliche Neubewertung des Alterns: Altern sei ›traditionell‹ erstens von jeher als eine Form des Verfalls verstanden worden, der der Einzelne zweitens machtlos gegenüberstünde. Beides kehren Rowe und Kahn (1997) um, indem sie den Alterns- mit dem Leistungsbegriff verschränken. Altern könne positiv gewendet werden und erfolgreich sein, wenn der Einzelne den eigenen Alternsprozess nur selbst gestaltet und den Verfall hinauszögert:

»Successful aging is multidimensional, encompassing the avoidance of disease and disability, the maintenance of high physical and cognitive function, and sustained engagement in social and productive activities. Research has identified factors predictive of success in these critical domains. The stage is set for intervention studies to enhance the proportion of our population aging successfully.« (ibid. 433)

Mit seiner Vorstellung von der »Intervention« wendet sich das »Successful Aging« letztlich gegen eine Medikalisierung des Alterns, die es primär mit Gebrechen, Krankheit oder Behinderung assoziiert. Die Leistung des Alternenden, um die die neue Gerontologie kreist, liegt darin, diesen Verfallsprozess aufzuhalten oder zumindest hinauszuzögern. Der Einzelne ist damit der Meister des eigenen Alternsprozesses. Für diese Vorstellung von ›Meisterschaft‹ sind Rowe und Kahn (1998) oft kritisiert worden, blenden sie doch die Rolle ökonomischer und sozialer Vorbedingungen aus: Den Marathon hin zum Ziel des hohen oder höchsten Alters, so die Kritiker:innen, beginnen eben nicht alle von der selben Startlinie aus; auf die Fähigkeit, dem kalendarischen Alter zu trotzen, können soziale Herkunft und andere Randbedingungen durchaus einen erheblichen Einfluss haben. Vergleichbar der schulischen Bewertung geht das Konzept des ›Successful Aging‹ aber davon aus, dass von diesen Randbedingungen, die erfolgreiches Altern ermöglichen, abgesehen werden kann. Damit werden der Leistungsgedanke und die Wettbewerbsidee auf einen Prozess übertragen, der bis dahin nur als eine biologische Gegebenheit verstanden worden war.

Eine weitere Parallele zur Schule ist diese: Erfolgt dort ein klasseninterner Vergleich von Schüler:innen durch Lehrkräfte, kommt es hier zu einem auf Alterskohorten bezogenen, ›klasseninternen‹ Vergleich von Älteren durch Mediziner:innen. Als Bewertende stehen diese, so wird zumindest von Rowe und Kahn (1998) impliziert, vollkommen außerhalb des Prozesses der Leistungsbewertung.

Zu den Gallionsfiguren des ›Successful Aging‹ werden, bald nach dem Erscheinen des Buches die Hundertjährigen. Es sind ihre Erzählungen in Interviews oder Autobiographien, die auf Buchmärkten weltweit Konjunktur haben (Banerjee 2023); sie werden auch für die Medizin zu Schlüsselfiguren der Erforschung extremer Langlebigkeit (Yashin et al. 2000). Auch hier ist das kalendarische Alter Ausgangspunkt für die Verschränkung zwischen Alter und Leistung; es gilt letztlich, den Tod hinauszuzögern. Nicht von ungefähr gibt R. Waldo McBurney, der hundertjährige Marathonläufer, seiner Autobiographie den Untertitel »My First 100 Years« (McBurney 2004).

Der Begriff des ›Successful Aging‹ wirft die Frage auf, wie dieser ›Erfolg‹ eigentlich gemessen werden kann und in welchen Vergleichen er bestimmt werden soll: Im Verhältnis zu wem genau altert man mehr oder weniger erfolgreich? Der Bereich des Sports schafft hier Abhilfe. Beim sportlichen Wettkampf lässt sich die Leistung älterer Athlet:innen direkt beobachten, messen und prämiieren. In den USA sind die »National Senior Games«, im Deutschen oft mit ›Seniorenolympiade‹ übersetzt, längst zu einem Sportereignis mit Millionenpublikum avanciert. Im Zentrum der Berichterstattung steht die Erbringung von Höchstleistungen einer ganz besonderen Art: die Idee, dass man dem kalendarischen Alter trotzen oder buchstäblich davonlaufen kann. Wie McBurney, der Marathonläufer, es beschreibt, »My national and international competition started when I was 88, and [had shifted] from distance running to shorter distances and race walking« (36). Hier wird der Alternsprozess als eine Veränderung, die nach wie vor mit der Verminderung körperlicher Leistungsfähigkeit einhergeht, zwar angesprochen, aber der hochaltrige Leistungsträger wechselt kurzerhand die Disziplin; anstatt zu sprinten, ›walked‹ er nun.

Hier wäre durchaus eine alternative Lesart denkbar, die die Verschränkung des Alterns- mit dem Leistungsbegriff weniger zentral gemacht hätte: Denn dass Sporttreiben im Alter gewinnbringend ist, dass man durch körperliche Fitness auch die eigene geistige ›Lebendigkeit‹ und Kreativität befördern kann, ist eine Ansicht, die nicht nur von Mediziner:innen und Physiotherapeut:innen geteilt wird. Bezeichnend ist hier aber, dass der hochbetagte Marathonläufer R. Waldo McBurney eben nicht einfach nur Sport treibt, um auch im Alter noch beweglich zu bleiben; die Leistung entsteht erst in der Rahmung des Marathonlaufs und der Seniorenolympiade. So mag man hier auch anführen, dass es kein Zufall ist, dass der Diskurs des Successful Aging seinen Ursprung in den USA hat, einer der am stärksten meritokratisch geprägten westlichen Gesellschaften. Der Alternsbegriff ist in diesem Sinne ebenso kulturgebunden wie die Verschränkung des Alterns- mit dem Leistungsbegriff, die in den USA sehr viel stärker ausgeprägt ist, als dies etwa in Deutschland der Fall ist (Lamb 2017). Dass Altern überhaupt mit Leistung in Verbindung

gebracht wird, steht in engem Zusammenhang mit dem Alters- und dem Leistungsverständnis westlicher Gesellschaften, während in nicht-westlichen Gesellschaften weiterhin die Idee der Lebensleistung (*merit*) im Zentrum von Altersvorstellungen steht (Selin 2022).

Während die Leistung von Schüler:innen primär durch andere optimiert werden sollen, kann das ›Successful Aging‹ vor allem als eine Form der fortwährenden Selbstoptimierung verstanden werden. Der Diskurs des ›Successful Aging‹ wird hier in die Praxis der Lebensführung übersetzt. Die Selbstdarstellung der Best Agers in Autobiographien oder Social Media-Profilen entspricht einem permanenten Self-Tracking, bei dem das Selbstnarrativ der Smart Watch entspricht. Die Selbstbeschreibungen sind voll von Metriken des Selbst (Reichardt und Schober 2020); in ihnen geben die hochaltrigen Leistungsträger detailliert Auskunft über ihre Lebensweise, von ihrem Schlafzyklus über ihre Ernährungsgewohnheiten. Die Selbstoptimierung des Einzelnen entlastet den Staatshaushalt und die Krankenkassen (Lemke 2013); die Annahme, der einzelne trage für sein Schicksal die alleinige Verantwortung, lässt soziale Sicherungssysteme zweitrangig erscheinen (Brown 2015).

Der Idee eines meritokratischen Alternsbildes mag es auch geschuldet sein, dass Erfolg ökonomisiert werden kann bzw. sogar muss. Während auf dem Gebiet des Sports Erfolg messbar ist, ist dies für den Bereich des Lifestyles nicht unmittelbar der Fall. Bezeichnend für die Verbindung von Altern und Leistung ist es auch hier, dass die Leistung nicht nur erbracht, sondern von anderen bezeugt werden muss; durch dieses Bezeugen oder »Liken« wird sie erst zur Leistung.<sup>2</sup> Damit wiederum werden die Social Media zu einem Raum, in der »Successful Aging« auch im außersportlichen Bereich vollzogen werden kann. Es ist bezeichnend, dass die Ikonen des ›Successful Aging‹ – von der Modedesignerin Iris Apfel, die ihre Modelkarriere mit 98 Jahren begann, bis zu Baddie Winkle, dem ältesten Unterwäschemodell der Welt – aus den USA stammen. Auch hier ließen sich Alternativen zum ›Successful Aging‹ denken: Wenn man »junggeblieben« die Kleider seiner Tochter trägt oder mit 60 für sich das Bergwandern entdeckt, ließe sich das, ganz unabhängig vom Leistungsbegriff, auch als das Neugestalten des eigenen Lebens und die Neuerfindung der eigenen Identität fassen. Der Leistungsbegriff kommt dann zum

- 2 Die Erbringung von Leistung durch Lifestyle auf Social Media erfolgt keineswegs nur im Zusammenhang mit Best Agers, sondern trifft auch für viele andere Gruppen zu. Die »Clicks« und das »Liken« wiederum sind nicht nur soziales Kapital, sondern sie sind auch umsetzbar in ökonomisches Kapital, das messbar ist. Influencer und »Granfluencer« (ältere Influencer) können auf ihren Social Media Accounts für Produkte werben, etwa, indem sie bestimmte Marken tragen oder ihre Posts mit Produktwebsites verlinken; umgekehrt können Unternehmen ihnen dann kostenlos Produkte schicken, die dann in Social Media Posts beworben werden (Delbaere et al. 2021).

Tragen, wenn diese Neuerfindung vor einer breiten Öffentlichkeit vorgeführt wird und prämiert werden will; mehr noch: wenn sie von den älteren Leistungsträger:innen selbst auch ökonomisch gerahmt wird. Daher stößt das Modell des ›Successful Aging‹ gerade in den USA auf so breite Akzeptanz, da seine Bevölkerung habituell auf individualistisches Engagement und Erfolg setzt. In diesem Land, das bis heute nicht alle seine Bürger:innen krankenversichert und das über kein umfassendes soziales Sicherungssystem verfügt, propagieren nun Best Agers wie Iris Apfel und Baddie Winkle, »retirement is a fate worse than death« (Elkins 2018). Sie dehnen somit den biographischen Lebensabschnitt der Erwerbstätigkeit bis weit über das Erreichen des Pensionsalters hinaus aus.

Damit werden der Alterns- und der Leistungsgedanke systematisch verschränkt. Läge es nahe, zu denken, anders als beim Sport könne Leistung im täglichen Leben nicht gemessen werden, so schaffen die Best Agers Abhilfe, indem sie diese »Leistung durch Lifestyle« ebenfalls auch im ökonomischen Sinne messbar machen: durch »Clicks« auf social media-Plattformen wie TikTok und Instagram ebenso wie durch das Verkaufen von eigenem ›Merch‹ und dem Eingehen von Modelverträgen. Baddie Winkle, die dadurch »entdeckt« wurde, dass ihre Enkelin ein Foto von ihrer »heißen Oma« im Minirock auf ihr Instagram-Profil lud, ging mit weit über 80 Jahren einen Vertrag mit Smirnoff ein und posierte in Werbespots als »real baddass« Party Girl. Es reicht für die »Feststellung« des ›Successful Aging‹ eben nicht aus, zur Rente etwas dazuzuverdienen; es geht um den Modelvertrag, den man erst mit 98 abschließt.

Bemerkenswert bei der Idee des ›Successful Aging‹ ist es, dass nicht mehr die Lebensleistung (*merit*) im Vordergrund steht – ein Aspekt, der in der Auffassung hohen Alters doch naheliegend wäre. Vielmehr geht es auch beim Altern nun nicht mehr um *merit*, sondern um Achievement; dieses *Achievement* ist es auch, das symbolisch gewürdigt und wie eine Art ›Orden‹ verliehen wird. Hier lohnt es sich auch, den Aus- bzw. Vorführungsbegriff zu berücksichtigen. In der Schule und der Fabrik wird Leistung ausgeführt, beim erfolgreichen Altern wird sie vorgeführt. Beim Successful Aging werden die Medien selbst zu einer Institution, in der die Alterns- (und nicht die Lebens-)leistung aufgeführt, bestaunt und prämiert werden kann. Die mediale Darstellung trägt damit selbst dazu bei, dass die Überführung von Lebensleistung in Successful Aging vollzogen werden kann.

Beim sportlichen Wettkampf ebenso wie bei Lifestyle-Performances handelt es sich um eine Aufführung, der ein Publikum beiwohnen muss: Eine Aufführung von ›Successful Aging‹ im eigenen Wohnzimmer oder nur vor der eigenen Familie wäre wenig zielführend; unterstellt wird hier, dass es ein Publikum sein muss, das – analog zur Schule – »objektiv« urteilt und sich nicht etwa von seiner Sympathie für die hochaltrigen Leistungsträger:innen leiten lässt. Damit stellen die ›Likes‹ der Profile von

Best Agern auf Social Media eine Form der Leistungsmessung dar, die als solche aber gar nicht taugt, weil sie – anders als beim Sport – nicht auf messbaren Kriterien beruht.

Anders als in der Schule kann man in der Annahme von ›Successful Aging‹ nur zwischen Erfolg und Misserfolg unterscheiden; es gibt keine Zensuren, und erfolgreiches Altern kann nicht auf einer Skala gemessen werden; überhaupt liegen keine messbaren Kriterien vor. Bemerkenswert an dem Diskurs über erfolgreiches Altern ist gleichzeitig, dass er implizit mit einer Gegenfolie operiert, die jedoch nie offen ausgesprochen oder explizit beschrieben wird. Denn die Bezeichnung ›erfolglosen‹ Alterns scheint sich aus ethischen Gründen zu verbieten; noch dazu ist auch hier unklar, worin sich ein solcher Misserfolg niederschlagen würde oder wie er gemessen werden könnte. Hierin liegt die Paradoxie des Diskurses um ›Successful Aging‹: Denn wenn keine cut-off-Linie existiert, anhand derer messbar wäre, wo erfolgreiches Altern aufhört und in erfolgloses Altern umschlägt, wird letztlich die Idee der Leistungserbringung selbst ad absurdum geführt.

Im Schatten der Idee des Successful Aging stehen all diejenigen, die den Marathon nicht laufen, an der Seniorenolympiade nicht teilnehmen und die eigenen Fotos nicht ins Netz stellen können oder wollen. Damit, auch wenn diese Kampfansage nie offen ausgesprochen wird, positioniert sich das Successful Aging in direkter Opposition zu körperlicher und mentaler Einschränkung oder Behinderung. Rowe und Kahn (1997) sprechen in der oben zitierten Passage nicht von ungefähr von der erfolgreichen »Vermeidung von Krankheit und Behinderung« (»avoidance of disease and disability«, 433). In all diesen Aspekten könnte die Vorstellung vom Successful Aging anfechtbarer nicht sein; eine Problematik, die nicht zuletzt in der Subsumtion des Alters- unter den Leistungsbegriff begründet ist.

In einer Zeit des demographischen Wandels hin zu »alternden Gesellschaften« stellt das gesellschaftliche Segment der Älteren die am stärksten wachsende Bevölkerungsgruppe dar. Es mag nicht zuletzt dieser demographischen Dynamik geschuldet sein, dass es jetzt zu einer Ausdifferenzierung *innerhalb* der Kategorie der Älteren kommt. Dass diese Ausdifferenzierung sich im Zusammenspiel mit dem Leistungsbegriff vollzieht, ist ein weiterer Beleg für die Konjunktur des Leistungsbegriffs in modernen westlichen Gesellschaften.

## 5. Fazit und Ausblick

Die vier hier diskutierten Fälle haben verdeutlicht, wie Leistung über einen Zeitraum von mehr als 100 Jahren zu einer gesellschaftlichen Leitvorstellung wurde. Zugleich hat die Untersuchung der Arbeitsphysiologie um 1900, eines Industriebetriebs in den 1920er und 1930er Jahren,

der neueren Forschung zu Leistungsbewertung in der Schule sowie des Diskurses und der Praxis des *Successful Aging* zu Beginn des 21. Jahrhunderts gezeigt, dass Leistung als Konzept und in ihrer Praxis eine variable Differenzierungsform ist. Mit der Gegenüberstellung historischer und gegenwärtiger, wissenschaftlicher und vernakularer Konzeptionen von Leistung haben wir die Vielstimmigkeit und soziale Wirkmächtigkeit des Leistungsgedankens in der Gegenwartsgesellschaft an exemplarischen Konstellationen anschaulich machen können. Auch wenn unsere Auswahl von Fällen natürlich hochselektiv ist, lassen sich aus ihrer Kontrastierung grundlegende Beobachtungen über die historische Entwicklung und die allgemeinen Mechanismen der Leistungsdifferenzierung formulieren.

Mit Blick auf das 20. Jahrhundert ist zunächst die zentrale Bedeutung von praxisorientierten Wissenschaften, wie der (Arbeits-)Physiologie, der Bildungsforschung oder der Medizin, für die Herausbildung von modernen Leistungsbegriffen zu konstatieren. Mit ihren Ideen und Metaphern, Techniken und Wissensbeständen nahmen diese Wissenschaften Einfluss auf politische und bürokratische Prozesse bis hin zu alltäglichen Sinnwelten. Sie erforschten nicht nur das, was sie Leistung nannten, in verschiedenen sozialen Feldern, sie waren darüber hinaus bestrebt, ihre Vorstellung von Leistung in Arbeit, Schule und Lebensführung im Alter normativ durchzusetzen. Um die Wende zum 20. Jahrhundert trachtete die wissenschaftliche und industrielle Auseinandersetzung mit menschlicher Tätigkeit danach, Leistung als quasi-physikalische Größe zu etablieren und fundierte damit die soziale Realität moderner, auf Arbeitsleistung bezogener Gesellschaften. Auch wenn die sozialwissenschaftliche Forschung schon zu Beginn des 20. Jahrhunderts die Konstruktivität der Schulleistung herausstellte, wurde Leistung damit weder dekonstruiert noch entmythologisiert. Vielmehr verdeutlicht die Auseinandersetzung um die Frage, wie schulisches Wissen oder erfolgreiches Altern zu bewerten und damit als Leistung zu markieren ist, die Wirksamkeit der Vorstellung einer objektiv messbaren, von den Bewerteten selbst hervorgebrachten Leistung.

Ferner wurde deutlich, dass der moderne Leistungsbegriff seit seiner Erfindung ein Objekt ist, das in seinem konkreten Sinngehalt und seinen Anwendungsformen umstritten ist und verhandelt wird (Gallie 1964). Neben der Spannung von Objektivierung und Konstruktivität, hat dieser Beitrag auch auf der Beobachtungsebene eine grundlegende Spannung konstatiert, die einerseits zwischen Individuum und Kollektiv verläuft und andererseits zwischen unterschiedlichen Zeitlichkeiten. Untersuchte die frühe (Arbeits-)Physiologie Leistung punktuell in zeitlich begrenzten Reihenversuchen, um sie als anthropologische Grundkonstante zu ermitteln, diente die Leistungserhebung im Industriebetrieb, in der Schule und dann in der sportlichen und beruflichen Aktivität von Senior:innen

als Markierung von Individuen, deren Leistung sowohl im Vergleich mit anderen durch eine Rangplatzierung als auch im Vergleich mit sich selbst als individuelle Leistungsentwicklung im (beruflichen, schulischen, altersbedingten) Lebenslauf dargestellt werden kann. Dies konnte bis hin zur Prognose zukünftiger Leistung und einer zu erwartenden Leistungsfähigkeit reichen.

Leistungsdifferenzierung kann in verschiedenen Sinnschichten und unterschiedlichen Aggregatzuständen (Hirschauer/Nübling 2021: 58) erscheinen. Bei den dargelegten Fällen fällt auf, dass eine Objektivierung und Verfestigung der Leistungsdifferenz neben Wertprädikaten oft in Form von Zahlen geschieht. An die Stelle einer Arbeit, Tätigkeit oder Praxis tritt ein Wert, der die Bemessung und Bewertung als Leistung zusammenfasst, dokumentiert und vergleichbar macht. Unabhängig von seiner messtheoretischen Relevanz und Güte und seiner zuweilen pragmatischen Vermischung, insinuiert er eine quasi-naturwissenschaftliche Objektivität des Urteils: Die (Arbeits-)Physiologie dokumentierte die Befunde ihrer einfachen Experimente durch Zahlen, der Industriebetrieb Bat'a verwendete Tabellen als etablierte Form der Lohndarstellung und auch die Schule kommuniziert, trotz schriftsprachlich verfügbarer Prädikate, ihre Beurteilung oft numerisch. Beim *Successful Aging* steht allein das hohe, in einer Zahl dokumentierte kalendarische Alter unter der Voraussetzung für eine Leistung, dass der- oder diejenige noch aktiv sein kann. Eine solche numerische Dokumentation ist eine denkbar einfache Methode: ohne weitere Auseinandersetzung mit dem Sachverhalt ist sie sofort verständlich und stark vergleichsaffin. Sie reduziert Komplexität, jedoch beruht die Hervorbringung dieser Zahlenwerte auf historisch innovativen Techniken bzw. Technologien des Experimentierens und des ökonomischen Rechnens und Verwaltens; ferner haben wir kritische Diskurse über Bewertungsmaßstäbe und -kriterien sowie soziale Kämpfe etwa um die Neubewertung des Alterns thematisiert.

Wir haben also die Grundannahmen einer objektiv feststellbaren Leistungsbegriffs hinterfragt. Im Hinblick auf die Schule zeigen wir, dass Leistung durch Lehrende nicht einfach nur ›gemessen‹ oder festgestellt, sondern erst durch ihre Wertzuschreibung hervorgebracht wird; und dass die Bewertenden in der Benotung ihrer Schüler:innen letztlich immer auch sich selbst bewerten. Die Idee der Optimierung von Leistung wiederum verbindet die moderne (Arbeits-)Physiologie, die Fabrikarbeit und die Idee des ›Successful Aging‹. In allen drei Fällen geht es immer auch um die Grenzen der Leistungsfähigkeit. Kraepelin schickte sich an, zu messen, wann körperliche Ermüdung eintrat und wie einer solchen Ermüdung entgegenzuwirken sei; Bat'a griff in die Lebensführung der Arbeitenden mit dem Ziel ein, ihre Arbeitsfähigkeit und -kraft fortwährend zu optimieren und unterzog sie deshalb einer ständigen Überwachung und Bewertung. Beim ›Successful Aging‹ greifen Selbst- und

Fremdüberwachung ineinander: Der Leistungsmessung durch Preisrichter und Zeituhren (bei der Seniorenolympiade) entspricht das permanente Self-Tracking durch den hochaltrigen Leistungsträger selbst. Dennoch ›kippt‹ im Fall des ›Successful Aging‹ das Bild bzw. die Technik der Leistungsmessung: Denn hier ist nicht das Alter selbst die Leistung (die Jahreszahl, der 100. Geburtstag), sondern diejenigen Handlungen, die man trotz des hohen Alters noch vollziehen kann. Das Alter wird hier analog zur Höhe der Hürde beim Sport verstanden, die es zu überspringen gilt. Dieses ›Überspringen‹ aber ist beim ›Successful Aging‹ sehr viel weniger klar als bei den anderen hier diskutierten Fällen. Dies wiederum kann als ein Beispiel für die Wirkmächtigkeit Leistungsgedankens verstanden werden: Die Semantik der Leistung bleibt auch dort wirksam, wo klare, messbare Kriterien darüber fehlen, was die Leistung eigentlich ausmachen soll.

Die Kontrastierung unserer Fälle zeigt auch andere Facetten des Leistungsbegriffs. So sind Leistungsdefinition und Leistungsmessung eng mit der Organisation verbunden, die sie rahmt. Die Organisation steht als ein Garant dafür, dass die Leistungsmessung legitim, wünschenswert und kriteriengeleitet ist. Sie absorbiert die Kontingenz der Leistungs differenzierung: Natürlich bewerten Lehrerinnen Gleiches verschieden, aber die Schule als Organisation verbürgt einen Objektivierungsprozess. Natürlich lässt sich die Leistung von Arbeitenden verschieden messen, einstufen und entlohnen, aber die (Arbeits-)Physiologie verspricht ein wissenschaftliches Fundament. Auch die Schuhfabrik als Organisation erhebt und verbürgt den Anspruch, die Leistung ihrer Angestellten zu objektivieren und zu ›ranken‹. Für das ›Successful Aging‹ wiederum dient in der neuen Gerontologie zunächst die Medizin als ein Feld, innerhalb dessen Erfolg nach klar definierten Kriterien (Abwesenheit von Krankheit im hohen Alter) gemessen wird, dann im Feld des Sports die Seniorenolympiade (National Senior Games) als eine Organisation, die über die Leistungsmessung und -bewertung ›wacht‹ und für ihre Richtigkeit bürgt. Beim erfolgreichen Altern als Lifestyle findet sich anstelle einer Organisation dagegen der Referenzrahmen des Publikums, dessen »Likes« als eine Art von Währung fungieren; eine übergeordnete, kriteriengeleitete Legitimierung der Bewertenden fehlt oder wird allenfalls durch bestimmte marktwirtschaftliche Grundprinzipien eines »Perform – or else« (McKenzie 2001) kompensiert. Hier franst der Leistungsbegriff konzeptuell aus, bleibt aber weiterhin ein zentraler Referenzpunkt.

In all diesen unterschiedlichen Feldern liegt die machtvolle Wirkung der Leistungsidee im Gleichheitsversprechen und dem emanzipatorischen Anspruch, der mit ihm verbunden ist. Die Leistungs differenzierung sollte andere Formen der Humandifferenzierung wie die nach familiärer Herkunft, Geschlecht, Ethnizität, sozialer Klasse oder Hautfarbe, allesamt ignorierbar machen, sobald nur die Leistung ›zählt‹. Es geht

dabei um legitime Stratifikation in primär sachlich differenzierten Gesellschaften, die andere Formen von Humandifferenzierung nach Möglichkeit abstreifen wollen. Auch aus diesem Grund kommt es regelmäßig zum Skandal, sobald Fälle von Manipulation, Bevorzugung oder Bestechung im Kontext von Leistungsvergleichen aufgedeckt werden. So etwa beim Doping oder der Wettmanipulation im Sport, aber auch in der Hochschulbildung, z.B. bei Plagiatsfällen oder als sich vor einigen Jahren herausstellte, dass die Leistungskriterien bei der Auswahl von Studierenden an Elite-Universitäten wie Yale oder Stanford durch ein ausgeklügeltes System von Spenden durch reiche Eltern umgangen wurden (Sandel 2020: 7–10).

All diejenigen, die an der Herstellung und Objektivierung von Leistung beteiligt sind, verpflichten sich mithin einer Idee, deren symbolischen Überschüsse sie ausblenden, die aber als Fiktion machtvoll und sozial wirksam ist. Die (Arbeits-)Physiologie untersucht im Zuge der Industrialisierung, wie Menschen als Arbeitskräfte optimiert werden können; die Fabrikarbeiterin glaubt daran, sie könne allein durch Leistung und harte Arbeit in der Hierarchie aufsteigen; jede Schülerin hofft, ihre Antworten, Klassenarbeiten etc. würden mit den gleichen Maßstäben bewertet wie die der Mitschüler:innen, die sozial besser gestellt sind oder phänotypisch mehr dem Lehrpersonal entsprechen; und jeder einzelne hofft, er könnte ungeachtet seiner ökonomischen Situation oder mangelnder Krankenversicherung gesund und im Vollbesitz seiner geistigen Kräfte 100 Jahre alt werden.

Damit ist der Glaube an Leistung und die eigene Leistungsfähigkeit in all diesen unterschiedlichen Feldern eine ›willing suspension of disbelief‹: Man einigt sich darauf, (immer wieder) zu vergessen, wie sehr Leistung und ihre Bewertung kontingent sind und bleiben. Damit liegt Leistung im Goodwill derjenigen begründet, die weiterhin und allen offensichtlichen Widersprüchen zum Trotz an sie glauben und ihre Ratifizierung praktizieren.

## Literatur

- Banerjee, M. (2023): *Centenarians' Autobiographies. Age, Life Writing and the Enigma of Extreme Longevity*, Berlin: DeGruyter.
- Birkel, P./Birkel, C. (2002): »Wie einigt sich Lehrer bei der Aufsatzbeurteilung? Eine Replikationsstudie zur Untersuchung von R. Weiss«, *Psychologie in Erziehung und Unterricht* 49: 219–224.
- Bourdieu, P. (1984): *Die feinen Unterschiede. Kritik der gesellschaftlichen Urteilskraft*, Frankfurt am Main: Suhrkamp.
- Bourdieu, P. (1989): *La noblesse d'État. Grandes écoles et esprit de corps*, Paris: Minuit.

- Breidenstein, G. (2023): »Prüfungswissen. Zur Formatierung schulischen Wissens durch Prüfungspraktiken«, *Zeitschrift für Pädagogik* 69(5): 568–583.
- Brookhart, S. M. (1993): »Teachers' grading practices: Meaning and values«, *Journal of Educational Measurement* 30(2): 123–142.
- Brookhart, S. M. (2004): »Classroom Assessment: Tensions and Intersections in Theory and Practice«, *Teachers College Record* 106(3): 429–458.
- Brown, W. (2015): *Undoing the Demos. Neoliberalism's Stealth Revolution*, New York: Zone Books.
- Čipera, D. (o.J.): *Personalakte zu Dominik Čipera, Státní okresní archiv Zlín, fond Baťa II/2, kart. 1021, inv. č. 13, č. 25*.
- Conrad, S. (2017): »Nothing is the way it should be: Global Transformations of the Time Regime in the Nineteenth Century«, *Modern Intellectual History* 15(3): 821–848.
- Cook, E. (2017): *The pricing of progress: Economic indicators and the capitalization of American life*, Cambridge, MA: Harvard University Press.
- Delbaere, M./Michael, B./Phillips, B. J. (2021): »Social media influencers. A route to brand engagement for their followers«, *Psychology & Marketing* 38(1): 101–112.
- Distelhorst, L. (2014): *Leistung. Das Endstadium der Ideologie*, Bielefeld: transcript.
- Doshal, Z. A. (2021): *In the Kingdom of Shoes. Bata, Zlin, Globalization, 1894–1945*, Toronto University Press.
- Elkins, K. (2018): »96-year-old Iris Apfel: »For me, retirement is a fate worse than death.«, in: *CNBC Make it*, 24.05.2018, <https://www.cnbc.com/2018/04/24/why-96-year-old-iris-afpel-is-never-going-to-retire.html> (letzter Zugriff am 24. Juli 2024).
- Etzemüller, T. (2009): »Social Engineering als Verhaltenslehre des kühlen Kopfes. Eine einleitende Skizze«, in: ders., *Die Ordnung der Moderne. Social engineering im 20. Jahrhundert*, Bielefeld: transcript, 11–39.
- Feindt, G. (2017): »Eine »ideale Industriestadt« für »neue tschechische Menschen«. Baťa's Zlín zwischen Planung und Alltag, 1925–1945«, in: Feindt, G./Gišibl, B./Paulmann, J. (Hg.), *Kulturelle Souveränität. Politische Deutungs- und Handlungsmacht jenseits des Staates im 20. Jahrhundert*, Göttingen: Vandenhoeck & Ruprecht, 109–132.
- Filer, A. (Hg.) (2000): *Assessment: Social practice and social product*, London: Routledge.
- Filer, A./Pollard, A. (2000): *The Social World of Pupil Assessment. Processes and Contexts of Primary Schooling*, New York/London: Continuum.
- Gallie, W. B. (1964): »Essentially Contested Concepts«, in: ders., *Philosophy and the Historical Understanding*, London: Chatto & Windus, 157–191.
- Hegarty, P. (2013): *Gentlemen's Disagreement. Alfred Kinsey, Lewis Terman, and the Sexual Politics of Smart Men*, University of Chicago Press.
- Heintz, B. (2010): »Numerische Differenz. Überlegungen zu einer Soziologie des (quantitativen) Vergleichs«, *Zeitschrift für Soziologie* 39: 162–181.
- Hirschauer, S./Nübling, D. (2021): »Sinnschichten des Kulturellen und die

- Aggregatzustände der Sprache«, in: Dizdar, D. et al. (Hg.), *Humandifferenzierung Disziplinäre Perspektiven und empirische Sondierungen*, Weilerswist: Velbrück Wissenschaft, 58–83.
- Hoche, A./Binding, K. (1920): *Die Freigabe der Vernichtung lebensunwerten Lebens. Ihr Maß und ihre Form*, Hamburg: Meiner.
- Hochweber, J. (2010): *Was erfassen Mathematiknoten?*, Münster: Waxmann.
- Janda, B. (1932–34): *Encyklopedie výkonnosti*, 3 Bde., Praha: Sfinx.
- Jürgens, E./Sacher, W. (2008): *Leistungserziehung und Pädagogische Diagnostik in der Schule: Grundlagen und Anregungen für die Praxis*, Stuttgart: Kohlhammer.
- Kalthoff, H. (1996): »Das Zensurenpanoptikum. Eine ethnographische Studie zur schulischen Bewertungspraxis«, *Zeitschrift für Soziologie* 25: 106–124.
- Kalthoff, H. (2017): »Kontingenz und Unterwerfung. Die organisierte Humanevaluation der Schule«, in: Hirschauer, S. (Hg.), *Un/doing Differences. Praktiken der Humandifferenzierung*, Weilerswist: Velbrück Wissenschaft, 259–284.
- Kaufman, B.E. (2010): *Hired Hands or Human Resources? Case Studies of Hrm Programs and Practices in Early American Industry*, Ithaca, NY: ILR Press.
- Keller, R. (1894): »Pädagogisch-psychometrische Studien«, *Biologisches Centralblatt*, Band XIV, 1, 2.
- Kraepelin, E. (1894): »Über geistige Arbeit«, *Neue Heidelberger Jahrbücher* IV, 1: 31–52.
- Lamb, C. (2017): *Successful Aging as a Contemporary Obsession: A Global Perspective*, Piscataway: Rutgers University Press.
- Lemke, T. (2013): *Biopolitik zur Einführung*, Hamburg: Junius.
- Luhmann, N. (2002): *Die Erziehung der Gesellschaft*, Frankfurt am Main: Suhrkamp.
- Luks, T. (2010): *Der Betrieb als Ort der Moderne. Zur Geschichte von Industriearbeit, Ordnungsdenken und Social Engineering im 20. Jahrhundert*, Bielefeld: transcript.
- McBurney, R.W. (2004): *My First 100 Years! A Look Back from the Finish Line*, Leawood: Leathers Publishing.
- McKenzie J. (2001): *Perform or Else: From Discipline to Performance*, London: Routledge.
- Meier, F./Peetz, T./Waibel, D. (2016): »Bewertungskonstellationen. Theoretische Überlegungen zur Soziologie der Bewertung«, *Berliner Journal für Soziologie* 26: 307–328.
- Offe, C. (1970): *Leistungsprinzip und industrielle Arbeit. Mechanismen der Statusverteilung in Arbeitsorganisationen der industriellen »Leistungsgeellschaft«*, Frankfurt am Main: Europäische Verlagsanstalt.
- Oggle, V. (2015): *The Global Transformation of Time. 1870–1950*, Cambridge, Mass: Harvard University Press.
- Parsons, T. (1991[1937]): *The Social System*, London: Routledge.
- Patzel-Mattern, K. (2010): *Ökonomische Effizienz und gesellschaftlicher*

- Ausgleich. Die industrielle Psychotechnik in der Weimarer Republik*, Stuttgart: Steiner.
- Pryor, J./Torrance, H. (2000): »Questioning the Three Bears. The social construction of class-room assessment«, in Filer, A. (Hg.), *Assessment: Social practice and social product*, London: Routledge, 110–128.
- Rabenstein, K./Reh, S./Ricken, N./Idel, T. (2013): »Ethnographie pädagogischer Differenzordnungen. Methodologische Probleme einer ethnographischen Erforschung der sozial selektiven Herstellung von Schulerfolg im Unterricht«, *Zeitschrift für Pädagogik* 59(5): 668–690.
- Rabinbach, A. (1990): *The Human Motor. Energy, Fatigue, and the Origins of Modernity*, University of California Press.
- Raphael, L. (1996): »Die Verwissenschaftlichung des Sozialen als methodische und konzeptionelle Herausforderung für eine Sozialgeschichte des 20. Jahrhunderts«, *Geschichte und Gesellschaft* 22: 165–193.
- Reichardt, U./Schober, R. (Hrsg.) (2020): *Laboring Bodies and the Quantified Self*, Bielefeld: transcript.
- Reh, S./Ricken, N. (Hrsg.) (2018): *Leistung als Paradigma. Zur Entstehung und Transformation eines pädagogischen Konzepts*, Wiesbaden: Springer.
- Rowe, J./Kahn, R. (1998): *Successful Aging*, New York: Random House.
- Rowe, J./Kahn, R. (1997): »Successful Aging«, *The Gerontologist* vol. 37 no. 4: 433–440.
- Sandel, M.J. (2020): *The Tyranny of Merit. What's become of the Common Good?*, London: Allen Lane.
- Schreiber, H. (1899): »Gegen Prüfungen und Noten«, *Zeitschrift für Philosophie und Pädagogik* 6: 31–38.
- Selin, H. (Hg.) (2022): *Aging Across Cultures: Growing Old in the Non-Western World. Vol. 10*, Cham: Springer Nature.
- Shay, S.B. (2004): »The Assessment of Complex Performance: A Socially Situated Interpretive Act«, *Harvard Educational Review* 74: 307–329.
- Steinführer, A. (2002): »Stadt und Utopie. Das Experiment Zlín 1920–1938«, *Bohemia* 43: 33–73.
- Stiggins, R.J./Conklin, N.F. (1992): *In teachers' hands. Investigating the practices of classroom assessment*, Albany: SUNY Press.
- Südkamp, A./Möller, J./Pohlmann, B. (2008): »Der Simulierte Klassenraum. Eine Untersuchung zur diagnostischen Kompetenz«, *Zeitschrift für Pädagogische Psychologie* 22: 261–276.
- Taylor, F. W. (1911): *The Principles of Scientific Management*, New York: Harper and Brothers.
- Verheyen, N. (2018): *Die Erfindung der Leistung*, Berlin: Hanser.
- Verkuyten, M. (2000): »School marks and teachers' accountability to colleagues«, *Discourse Studies* 2: 452–472.
- Weischer, W. (2017): »Gesellschaftlicher und sozialstruktureller Wandel 1800–2000. Überlegungen zu einer praxeologischen Protheorie«, *Archiv für Sozialgeschichte* 57: 65–110.
- Wihstutz, B. (2021): »Leistung und Devianz um 1900. Über Performances als Praktiken der Humandifferenzierung«, in: Dizdar, D. et al. (Hg.),

- Humandifferenzierung Disziplinäre Perspektiven und empirische Sondierungen*, Weilerswist: Velbrück Wissenschaft, 230–259.
- Yashin, A. I./De Benedictis, G./Vaupel, J. W./Tan, Q./Andreev, K. F./Iachine, I. A.; Franceschi, C. (2000): »Genes and longevity: lessons from studies of centenarians«, *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 55(7): B319–B328.
- Zaborowski, K./Meier, M./Breidenstein, G. (2011): *Leistungsbewertung und Unterricht. Ethnographische Studien zur Bewertungspraxis in Gymnasium und Sekundarschule*, Wiesbaden: VS.
- Zlín (1932): »Ti, kteří vedou ve své práci«, *Zlín* 12.08.1932, 3.