

Testpsychologie und Diskriminierungsfreiheit – Fairness diagnostischer Prozesse aus intersektionaler Perspektive

Maré Kondruß & Mai-Anh Boger

Zusammenfassung: Der Wunsch nach einer diskriminierungsfreien Diagnostik in Schule und anderen pädagogischen Settings wird häufig an die Psychodiagnostik herangetragen: Durch objektive und faire Tests soll diese Möglichkeiten einer differenzsensiblen Diagnostik schaffen. Doch wie be- und verantwortet die Testpsychologie diese gestellten Ansprüche an Objektivität und Fairness vor dem Hintergrund sprachlicher und kultureller Diversität? Wie sind Versuche von „kulturfairen“ Testkonstruktionen motiviert und welche Rolle spielen die Kontexte, in denen sich die testenden und getesteten Subjekte befinden? Anhand dieser Problemfelder erörtert der Artikel Fairness in der Psychodiagnostik aus einer intersektionalen Perspektive. Dazu werden zunächst der Diskursstand der Psychometrie/Psychodiagnostik skizziert und bisherige Versuche der Problemlösung kritisch untersucht. Schließlich werden intersektionale Perspektiven auf Psychodiagnostik entwickelt und es wird gefolgert, dass Fairness nicht *im* Test, sondern in der *Relation* zwischen Test und involvierten Subjekten innerhalb von Kontexten verortet ist.

Schlüsselwörter: Psychodiagnostik, Testpsychologie, Fairness, Diskriminierung

Title: Test Psychology and Non-Discrimination – Fairness of Diagnostic Processes from an Intersectional Perspective

Summary: Oftentimes, diagnosticians in schools and other educational settings rely on the hope for objective and fair psychological tests to ensure discrimination-free testing. But how does test psychology account for these claims made for objectivity and fairness in a society that is culturally diverse? What motivates the construction of „culturally fair“ tests, and which role do contextual factors play? Along with these questions, the following article discusses fairness and equity in psychodiagnostics from an intersectional perspective. To do so, it first outlines the state of discourse in psychometrics/psychodiagnostics and critically examines previous attempts to solve the problem. Finally, it develops intersectional perspectives on psychodiagnostics. It concludes that fairness is not a subject of (only) test construction but of the relationship between tests and the subjects involved within cultural contexts.

Keywords: Psychodiagnostics, Test Psychology, Fairness, Discrimination

1 Problemaufriss: Kultur, Sprache und Migration in der Psychodiagnostik

„Die Entscheidungen im Bildungswesen sind oft von großer Tragweite. Also ist es doch besser, Testergebnisse als Grundlage der Entscheidung zu nehmen. Man kann auch bei späterem Mißerfolg immer noch sagen: ‚Aber damals zeigte das Testergebnis ganz klar‘“ (Lorenz 1990: 146).

Mit diesem ironischen Kommentar eröffnet Alfred Lorenz ein Spannungsfeld auf Basis der Beobachtung, dass die Hoffnung auf objektive Tests häufig aufgerufen wird, um Entlastung und Sicherheit in weitreichenden Entscheidungsprozessen zu gewinnen – ob im Rahmen der Leistungsdiagnostik oder auch in der Diagnostik im Kontext sonderpädagogischer Förderung. Angesprochen wird darin die legitimatorische Funktion von Tests sowie die Möglichkeit, diesen die ‚Schuld‘ daran zu geben, wenn in Bildungseinrichtungen etwas nicht so läuft wie es sollte. Mit Blick auf intersektionale Perspektiven schwingt darin auch mit, welche Strukturen und Praktiken die Reproduktion institutioneller Diskriminierung (Gomolla/Radtke 2009) mit zu verantworten haben. Statistiken zeigen dabei in ihrem Objektivitätsanspruch ein ambivalentes Gesicht: Einerseits können sie dazu verwendet werden, ebensolche Formen institutioneller Diskriminierung aufzudecken und sie messbar zu machen. Andererseits können Objektivitätsansprüche selbst in die Reproduktion von struktureller Diskriminierung verstrickt sein – so zum Beispiel durch die Naturalisierung impliziter und nicht-reflektierter Normalitätsvorstellungen in Testverfahren.

Intersektionalität wird in vorliegendem Artikel als ein plurales Bündel aus Theoriezugängen verschiedenster Provenienz und praktischen Zugängen verstanden, die darüber vereint werden, dass sie auf eine Kritik diskriminierender Verhältnisse mit Fokus auf intersektional invisibilisierte Personen, Konstellationen und Gruppen zielen (genauer in Abschnitt 2.2). Aus einer solchen diskriminierungskritischen Perspektive fragt man sich in der pädagogischen Praxis oft, wem man wie sehr vertrauen kann, etwas diskriminierungsfrei beurteilen zu können: *Auf wen oder was hoffen, wenn man auf eine diskriminierungsfreie Diagnostik zielt?* Setzt man auf Diskriminierungsfreiheit durch eine Sensibilisierung von Lehrkräften und anderen pädagogischen Fachkräften? Und/oder vertraut man eher auf Testverfahren, die als hinreichend objektiv, reliabel und valide gelten, und hofft, damit der subjektiven Willkür ein Ende zu setzen?

Intersektionalität ist zudem damit assoziiert, auf ebenjenes zu verweisen, das sich einer einfachen – wie man in der Statistik sagen würde – univariaten Kategorisierung entzieht, und auf komplexe Überkreuzungen und Interaktionen von Variablen zu verweisen, die mit *qualitativen* Sprüngen einhergehen (vgl. Walgenbach 2012). Implizit wird damit auch auf die mitunter in der Spur des Methodenstreits geführten Diskussionen zum Unterschied zwischen qualitativen (idiographischen) und standardisierten quantitativen (nomothetischen) diagnostischen Verfahren angespielt. Vorliegender Beitrag widmet sich daher der Frage, ob und wenn ja, wie Fairness über standardisierte diagnostische Verfahren gewährleistet werden kann. Der eingangs zitierte persiflierende Ausspruch verweist insgesamt auf drei Phänomene, die im vorliegenden Beitrag aus der Perspektive der Testpsychologie sowie aus einer intersektionalen Perspektive erörtert werden sollen.

Erstens wird im Zitat deutlich, dass es bei Diagnostik in der Praxis um Fragen der *Verantwortung* geht: *Wie werden Ansprüche an Objektivität und Fairness vor dem Hintergrund kultureller und sprachlicher Diversität in der Testpsychologie be- und verantwortet (2.1)? Und welche Verantwortung kommt den Diagnostizierenden in der Praxis zu (2.2)?*

Zweitens zeigt das Zitat, dass der Rückgriff auf diagnostische Instrumente und Verfahren *unterschiedlich motiviert* sein kann. Persifliert wird hier die auf den ersten Punkt bezogene Motivation, sich der Verantwortung durch Rekurs auf die Objektivität und Validität eines psychometrischen Verfahrens zu entziehen. Bezogen auf Diskriminierung kulminiert eine solche Verantwortungsverschiebung in der beruhigenden Vorstellung, dass nicht man selbst, sondern eben der Test diskriminierend war. *Wie sind demgegenüber die testtheoretischen*

Versuche, Fairness herzustellen, motiviert (3.1)? Und welche (weiteren) Motive finden sich in der Praxis der Durchführung und Interpretation dieser Verfahren (3.2)?

Drittens eröffnet das Zitat die Frage danach, wie diagnostische Prozesse und Ergebnisse durch den *Kontext* mitbeeinflusst werden. Der kritische Diskurs darüber kann sich auf zwei Ebenen entfalten. Betrachtet werden können einerseits kulturelle Kontexte (4.1) und andererseits Kultur als Kontext (4.2). *Was ist ‚Kultur‘ in der Testpsychologie und in der Intersektionalitätsforschung? (4.3)*

Nach Erörterung dieser drei Dimensionen — der ethischen, der motivationalen und der kontextuellen —, wird im Fazit eine mögliche Pointe einer intersektionalen Perspektive auf Diagnostik entfaltet (5).

2 Ethische Dimension: Verantwortung und Fairness

Aus testtheoretischer Perspektive ist die große Frage nach ‚Gerechtigkeit‘ *notwendigerweise* ausgeklammert, da Gerechtigkeit nicht test- bzw. verfahrensimmanent gewährleistet werden kann. Es wäre also ein Kategorienfehler, die Testdiagnostik dafür zu kritisieren, dass sie keine Gerechtigkeit herstellt. Im Zentrum stehen hier Konzepte der (prozeduralen) *Fairness*. Diese werden im folgenden Abschnitt (2.1) dargelegt, bevor aus intersektionaler Perspektive über Diskriminierungsfreiheit als Facette von Gerechtigkeit eingegangen wird (2.2).

2.1 Konzeptionen von Fairness in der Testpsychologie: Erkannte Probleme und Zielformulierung

Ausgangspunkt der selbstkritischen Anfragen und Überarbeitungen der Testdiagnostik sind zumeist Befunde der Unter- oder aber der Überdiagnostik im Intergruppenvergleich. So deuten beispielsweise groß angelegte epidemiologische Studien auf eine systematische Unterdiagnostik von ADHS bei Kindern in Familien mit Migrationshintergrund hin (Santos-Hövenner et al. 2019, Schlack et al. 2007). Interpretationen von statistischen Befunden dieser Art sind jedoch ungenügend begründbar, solange sie nicht mit einer Theorie bzw. einem komplexeren Modell zeigen können, dass von einer gleichen Häufigkeit bei allen Gruppen plausiblerweise ausgegangen werden kann (oder eben, dass dies nicht der Fall ist). Im Bereich Geschlecht liegen hierzu vereinzelt komplexere Theorien vor.¹ Für andere Heterogenitätsdimensionen – und so auch für die hier relevante Dimension kultureller Heterogenität – wurde dies bisher jedoch selten getan.

Auch die Vielfalt an Sprachen ging mit Anfragen an die Testtheorie einher. Im Zuge der Globalisierung wurde es eine geläufige Praxis, Tests in verschiedene Sprachen zu übersetzen. Dabei ist jedoch nicht garantiert, dass übersetzte Items in einer anderen Sprache eine äquivalente Bedeutung tragen. Es ist also weder gesichert, dass dasselbe Konstrukt gemessen wird, noch lassen sich folglich die Testwerte von Personen unterschiedlicher Sprache oder

1 Ein anschauliches Beispiel hierfür ist die Anorexie: So wäre es offensichtlich verkürzt, eine Unterdiagnostik von Anorexie bei Jungen/Männern anzunehmen, *nur* weil diese Diagnose für dieses Geschlecht seltener gestellt wird. Warum diese Störung bei Mädchen/Frauen häufiger auftritt, lässt sich geschlechtertheoretisch erörtern – und erst auf Basis dieser theoretischen Erörterungen lassen sich die statistischen Befunde sodann sinnhaft interpretieren.

verschiedener kultureller Herkunft fair vergleichen. So wurde zum Beispiel die häufig eingesetzte *Wechsler Intelligence Scale for Children* (WISC-IV, Wechsler 2003) in viele europäische Sprachen übersetzt, eine Vergleichbarkeit der verschiedenen Sprachversionen gilt jedoch nicht als gesichert (vgl. Kubinger 2011: 18 f.).

Vor dem Hintergrund dieser beiden Problemanzeigen wurde Test-Fairness im historischen Prozess zunehmend thematisiert und als ein weiteres Gütekriterium hochgehalten. Definiert ist das Gütekriterium der Test-Fairness laut Testkuratorium (1986: 360) als das „Ausmaß einer eventuell bestehenden systematischen Diskriminierung bestimmter Testpersonen, z. B. aufgrund ihrer ethnischen, soziokulturellen oder geschlechtsspezifischen Gruppenzugehörigkeit, bei der Abschätzung von Kriteriumswerten“. Als „fehlende systematische Benachteiligung“ (Arendasy et al. 2018: 42) umfasst Fairness nicht nur die inhaltliche Ebene des Messinstruments, sondern auch die prognostische Fairness und die Selektionsfairness.

Als Lösungsvorschläge für die genannten Probleme wurden sog. ‚kulturfreie‘ bzw. ‚kulturfaire‘ Tests konstruiert (vgl. Möbus 1983), auf die im nächsten Abschnitt genauer eingegangen wird (3.1). Hinter diesen verbirgt sich ein Verständnis von Test-Fairness im Sinne des „Identitätskonzepts“ (Wottawa/Amelang 1980: 200). Dieses besagt, dass ein fairer Test keine Unterschiede für die Zielgruppen zeigen und folglich auch nicht mit kulturellen Faktoren korrelieren darf (ebd.; vgl. auch Arvey 1972; Darlington 1971; Möbus 1983). In dieser Definition von Fairness taucht die oben benannte theoretische Lücke erneut auf: So lässt sich ohne ein theoretisches Modell nicht festlegen, ob gefundene Intergruppen-Unterschiede auf Verzerrungen oder auf Korrelationen mit kulturellen Faktoren zurückzuführen sind (ob also zum Beispiel das Konstrukt selbst ein westliches ist).

Das Identitätsprinzip liegt auch der verbreiteten Praxis zugrunde, bei der Fairness über Normtabellen hergestellt werden soll (vgl. Wottawa/Amelang 1980: 201 f.). Aufgrund theoretischer Überlegungen gibt es üblicherweise Normen für Alter oder Geschlecht, jedoch kaum Normen für soziale Schichten oder kulturelle Faktoren. A.a.O. wird diese Willkür anschaulich am Beispiel der Intelligenzdiagnostik erläutert:

„die eigentlich naheliegende Erklärung, daß eben 50-jährige ‚dümmer‘ sind als 30-jährige, wurde offensichtlich aus inhaltlichen Überlegungen heraus nicht akzeptiert (währenddem man sich an den größeren Mittelwertsdifferenzen von Angehörigen verschiedener Sozialschichten nie gestört hat)“ (ebd.: 201).

Nicht jeder im Intergruppenvergleich gefundene Unterschied wurde in der Testdiagnostik auf dieselbe Weise behandelt. In solchen Momenten zeigen sich die Einflüsse diskriminierender Stereotype und Diskurse auf Fragen der Testkonstruktion und deren Bewertung als (un-)fair.

Zu den kritischsten ihres Faches können jene gezählt werden, deren Definitionen von Testfairness auch test-exmanente Prüfkriterien einbeziehen:

„Als unfair gilt ein Test, wenn er in verschiedenen Gruppen oder Kulturen aufgrund diagnostisch irrelevanter Faktoren [...] unterschiedlich schwierig, trennscharf oder valide ist und diagnostische Entscheidungen [...] zur Benachteiligung bestimmter Gruppen führen“ (Schmitt/Maes/Seiler 2001: 88).

Diese Definition, die als Prüfkriterium auch (nicht-intendierte) diskriminierende Effekte in der realen Welt, in der diese Tests eingesetzt werden, miteinbezieht, hat jedoch keinen hegemonialen Status erreicht (genauer in 4.1).

Mit Blick auf die Teilfragestellung, wie die Testdiagnostik Ansprüche an Objektivität und Validität vor dem Hintergrund kultureller und sprachlicher Diversität verantwortet, lässt sich festhalten, dass das Entwerfen des Gütekriteriums der Test-Fairness zeigt, dass diese Frage durchaus ernstgenommen wurde. Die Sichtung dieser Entwürfe legt jedoch ebenso offen, dass

wir von diesem Ziel noch sehr weit entfernt sind (genauer in 3.1). Test-Fairness erweist sich für die meisten diagnostischen Verfahren als Desiderat und nicht als etwas, das als gegeben vorausgesetzt werden könnte.

2.2 Gerechtigkeit und Fairness im diagnostischen Prozess aus intersektionaler Perspektive

Bei intersektionalen Ansätzen handelt es sich nicht im strengen Sinne um ein wissenschaftliches Paradigma. Vielmehr wird unter diesem Zeichen ein sehr plurales Bündel an Theoriezugängen und Praktiken zusammengefasst, die sich gegen Diskriminierung richten und die dabei drei Momente akzentuieren (Boger 2022): Erstens fordern intersektionale Zugänge eine Komplexitätssteigerung ein. Durch den Verweis auf mehrfache und sich überkreuzende Benachteiligungen arbeiten sie heraus, dass unterkomplexe (uniaxiale, monokausale) Konzepte von Diskriminierung selbst Diskriminierungen hervorbringen (ex. s. Crenshaw 1989). Zweitens verdeutlichen sie dies, indem sie im Sinne der „intersectional invisibility“ (ex. s. Purdie-Greenaway/Eibach 2008) auf Konstellationen hinweisen, die durch diese unterkomplexen Vorstellungen unsichtbar gemacht werden. Drittens oszillieren sie dabei zwischen zwei Polen, die in öffentlichen Diskursen stark politisiert sind: Manche Vertreter:innen wenden den Verweis auf intersektionale Betroffenheiten so, dass sie in einem identitätspolitischen Modus Identitäten vervielfältigen. Andere wiederum erachten diesen Weg als aporetisch, insofern er das Problem, dass jede Repräsentation mit einem Ausschluss, einem Verfehlen einhergeht, nur vervielfältigt. Daher wird in dieser zweiten Stoßrichtung versucht, die Logik von Identität und Repräsentation selbst zu unterlaufen.

Als inhärent normatives und politisches Projekt zentriert die hier verfolgte intersektionale Perspektive Diskriminierungsfreiheit als Facette von Gerechtigkeit. In dieser Spur wird unter Intersektionalität demnach kein bestimmtes methodisches oder analytisches Verfahren verstanden, sondern es wird von intersectional (in)justice als normativem Fluchtpunkt ausgegangen, von dem aus eine Kritik der Reproduktion intersektionaler Diskriminierung in (pädagogischen) Praktiken vollzogen werden kann. Daher die hier formulierte Frage danach, welche Verantwortung den Diagnostizierenden in der Praxis zukommt.

Aus der Tatsache, dass Test-Fairness für die meisten diagnostischen Verfahren ein bisher uneingelöstes Desiderat ist, folgt aus intersektionaler Perspektive zweierlei. Erstens gilt: Der Einsatz testpsychologischer Verfahren ist keine neutrale Praxis. Die zweite These folgt aus dieser ersten: Die Praktizierenden tragen daher eine nicht-delegierbare Verantwortung für deren Einsatz. Fokussiert werden daher im Folgenden erstens Diskriminierungskritik als Verweis auf die *Nicht-Neutralität der vermeintlich objektiven Perspektive* und zweitens die *Nicht-Delegierbarkeit der Verantwortung*.

Beide Aspekte kommen in der Analyse der im Eingangszitat persiflierten Haltung zusammen. Wenn kontrafaktisch eine Objektivität behauptet und die (inter-)subjektive Dimension des diagnostischen Prozesses abgewertet (oder sogar ganz verleugnet und gestrichen) wird, entzieht sich das diagnostizierende Subjekt damit der Verantwortung. Es versucht gewissermaßen, diese an die Testpsychologie zu delegieren. Wie jedoch gezeigt werden konnte, behauptet die Testpsychologie gar nicht, diese Ansprüche an Objektivität und Fairness einzulösen. Beides sind Desiderate bzw. handlungsleitende Ideale innerhalb der dort verwendeten quantitativen Paradigmen, aber sie können nicht einfach als gegeben voraus-

gesetzt werden. Sobald aber die Nicht-Neutralität der vermeintlich objektiven Perspektive angenommen wird, verschwindet auch die Möglichkeit der Dialogverweigerung durch Rückzug auf die Position, man habe mit Hilfe des Tests objektive Erkenntnis erlangt. Insofern die diagnostizierenden Lehrkräfte/Fachkräfte keine Interpretationsgemeinschaft mit den Eltern, geschweige denn mit den Kindern bilden, kann also davon ausgegangen werden, dass sich der Anspruch auf Objektivität in der Praxis der dialogischen, also *inter-subjektiven* Validierung entgegenstellt. So wird in dieser Praxis nicht nur die Subjektivität der testenden Lehrkraft/Fachkraft gestrichen, sondern eben auch die des potentiell kritisierenden Anderen, dem durch die Ent-Subjektivierung auch das Gegenüber der kritischen Hinterfragung entzogen wird. Daher eröffnet bereits das Behaupten von Objektivität (und nicht erst das Behaupten von Fairness) ein ethisches Problem.

Die Intersektionalitätsforschung lädt dazu ein, sich diese Prozesse und Praktiken vom Standpunkt jener aus vorzustellen, für die Fairness nicht der Normalfall ist, den man selbstverständlich voraussetzt, und die das Behaupten von Objektivität historisch (kollektivbiographisch) und/oder in ihrer individuellen Biographie als Herrschaftspraxis erfahren haben. Von diesem Standpunkt aus ergeben sich zahlreiche weitere ethische Fragen: Mit dem Begriff des ‚white gaze/western gaze‘ kann zum Beispiel gefragt werden, inwiefern das diagnostische Verfahren selbst von einem betroffenen Subjekt als inquisitorisch bzw. gewaltförmig erlebt wird, gerade weil es als ‚objektiv‘ und objektivierend präsentiert wird (hooks 1984). In jedem Fall greift die Vorstellung, dass Objektivität Gerechtigkeit gewährleiste, zu kurz. Ein auf diesen Objektivitätsbegriff gestütztes Konzept von Fairness verkennt die hohe ethische Bedeutung des Sich-Verantwortens im inter-subjektiven Dialog. Dies ist eben nicht die ‚Schuld‘ der Testpsychologie, denn – genau dies ist die ethische Pointe – das Subjekt dieses inter-subjektiven Dialogs ist schließlich die praktizierende Lehrkraft/Fachkraft, die sich entscheiden und verantworten muss, ob sie sich hinter der Behauptung von Objektivität zurückzieht und als Mensch zum Verschwinden bringt, oder aber dem Kind ein pädagogisches Gegenüber – was immer bedeutet: ein (ver-)antwortendes Gegenüber – sein will. Dies gilt freilich für qualitative diagnostische Verfahren ebenso. Der Unterschied besteht lediglich in der im Eingangszitat persiflierten Häufigkeit, mit der in der Praxis der Versuchung nachgegangen wird, dies für scheinbar ‚objektive‘ standardisierte Verfahren zu verleugnen. Als nächstes soll daher die motivationale Ebene genauer untersucht werden.

3 Motivationale Dimension: Zur Differenz zwischen Erkenntnisinteresse und praktischen Interessen

Ein innerhalb der Testpsychologie maßgebender Ansatz, um das Problem der kulturellen Bias (vermeintlich) zu umgehen sind die sogenannten kulturfreien/kulturfairen Tests. Sie zeigen in ihren unterschiedlichen Konstruktionen verschiedene Versuche auf, Objektivität herzustellen, sind jedoch alle dadurch motiviert, die Konstrukte selbst weiterhin als kulturunabhängige verstehen zu wollen und genau dies durch kulturfreie/kulturfaire Tests zu affirmieren bzw. belegen zu wollen. Diese Motivation beeinflusst die Art und Weise, wie sich dem Problem prozeduraler Fairness aus disziplinärer Perspektive gewidmet wird (3.1). Die Motivationen und Interessen der Profession können daran nicht bruchlos anschließen (3.2).

3.1 Motivationen zur Problemlösung: Fairness durch „kulturfreie Tests“?

Bei den sog. kulturfairen Tests handelt es sich um eine heterogene Testgruppe: Kulturelle Verzerrungen betreffen verschiedene Testklassen wie Intelligenztests, neuropsychologische Testverfahren und Persönlichkeitsfragebögen (vgl. Bathje/Feiss 2014: 358). Auffällig hierbei ist, dass die Debatte um Kulturfreiheit und Kulturfairness überwiegend im Bereich der Intelligenz vorangetrieben wird, da für andere Testklassen nur wenige Versuche einer entsprechenden kulturfreien bzw. -fairen Testentwicklung unternommen werden. Die hier wiedergegebenen Ansätze, kulturelle Bias abzumildern, stammen daher alle aus dem Bereich der Intelligenztests. Zugleich erweist sich dieser Fokus auf Intelligenztests aus erziehungswissenschaftlicher Perspektive als vielfältig relevant, insofern diese sowohl im Bereich der Begabungsförderung als auch im sonderpädagogischen Bereich in der Praxis etabliert sind (Kottmann 2006). In beiden Feldern wurde auch angeführt, dass die mangelnde Fairness von Intelligenztests zur Reproduktion von Diskriminierungen beiträgt (z. B. im Kontext der Untertiagnostik von Hochbegabung bei Mädchen (vgl. Petersen 2013)).

Historisch betrachtet hat sich die Definition kulturfreier bzw. kulturfairer Tests innerhalb der Wissenschaft stetig gewandelt. Die ersten kulturfreien Konstruktionsbestrebungen wurden Ende der dreißiger Jahre unternommen: Raymond Cattell veröffentlichte 1940 den *Culture Fair Test*. Besonderen Aufschwung fand der Diskurs um kulturfreie Tests ab den 1960er Jahren im Rahmen der allgemeinen Kritik der Psychodiagnostik und insbesondere der Intelligenzdiagnostik, welche historisch darin verstrickt war, eine angebliche ‚genetische Unterlegenheit‘ bestimmter Gruppen zu ‚belegen‘ (Bathje/Feiss 2014; Benetka 1997).

Den zu Beginn der Debatte konstruierten Tests, unterstellte man kulturfrei zu sein. Cattell (1940: 165) versuchte zunächst, diese Kulturfreiheit durch die Verwendung eines „common ground of knowledge on which operations of reasoning could be performed“ zu gewährleisten, also durch den Einsatz von Testmaterial, welches Personen einer Kultur nicht vertrauter ist als Personen jeder anderen Kultur. Die postulierte Kulturfreiheit implizierte zudem eine universelle Einsetzbarkeit der Tests in jedem Kontext und jeder Population (vgl. Anastasi 1964: 27). Unter dieser Art von Test versteht man klassischerweise perzeptuelle Intelligenztests, also Tests, die möglichst kulturell nicht gebundenes, non-verbales Stimulusmaterial benutzen – wie z. B. einfache geometrische Objekte, zwischen denen abstrakte Relationen hergestellt werden sollen. Hierdurch sollten Verzerrungen durch Sprachfähigkeit oder kulturgebundene Inhalte vermindert werden (vgl. Cattell 1940). Ein weiteres prominentes Beispiel hierfür sind Ravens Progressive Matrizen von 1938 (Raven 1941).

Die postulierte Kultur- und Sprachfreiheit non-verbaler Tests wird im heutigen Fachdiskurs jedoch als nicht haltbar betrachtet. So erfordert z. B. das Lösen non-verbaler Testaufgaben zumeist stilles Verbalisieren (vgl. Kubinger 2011: 22). Beim non-verbalen Matrixtest müssen Regeln von links nach rechts abgeleitet werden, Testergebnisse können also durch die gewohnte Leserichtung (links nach rechts vs. rechts nach links) verzerrt sein (vgl. Röder et al. 2021: 24). Auch in empirischen Untersuchungen zeigten sich Verzerrungen durch non-verbale Tests (vgl. z. B. Dutton et al. 2018; Rosselli/Ardila 2003). Auf theoretischer Ebene wird dies auch dadurch leicht einsehbar, dass sich Kultur nicht auf Sprache reduzieren lässt. Das Vor- oder Nicht-Sprachliche ist nicht automatisch oder zwangsläufig ‚kulturfrei‘ oder ‚universal‘.

Aus diesen Gründen wurde das Konzept der Kulturfreiheit in den 1960ern – auch im Rahmen der Anerkennung der Interaktionen zwischen Anlage- und Umweltfaktoren – ver-

worfen und durch ein neues Konzept der „Kulturfairness“ ersetzt (Anastasi 1964: 26). Versucht wurde nun, zielgruppenspezifisch jeweils alle relevanten kulturellen Parameter zu eliminieren, in denen sich die getesteten Gruppen unterscheiden (ebd.). Bezüglich der Frage, wie die testtheoretischen Versuche, Fairness und Diskriminierungsfreiheit herzustellen, motiviert waren, lässt sich konkludierend festhalten, dass das Interesse der Rettung des Konstrukts und der Verteidigung der Universalität desselben galt. Dazu wurde zunächst versucht – in einer Metapher gesprochen – das Labor noch steriler zu machen, das Universale gewissermaßen zu destillieren. Implizit wird Kultur dabei als zu bereinigende Störvariable betrachtet. Diese Bereinigung geht mit einer Elimination der anthropologischen Dimension von Kultur einher: Der hier erträumte Test, so scheint es, soll so konsequent universalistisch sein, dass er rein hypothetisch auch für Affen und Aliens valide wäre.

3.2 Diskriminierungskritische Motivationen

Im Rahmen des disziplinären Denkens der Testpsychologie hat nicht das Wirksamkeitsinteresse, sondern das Erkenntnisinteresse Primat. Was aus den beschriebenen Universalisierungsversuchen in und für die Praxis folgt, muss daher separat erörtert werden. Im günstigsten Fall könnte aus diesen disziplinären Bewegungen eine diskriminierungsfreiere Praxis erwachsen. Dies ist jedoch weder gesichert, noch der Kernauftrag dieser Disziplin.

Im diskriminierungskritischen Diskurs wiederum lassen sich zwei gegenläufige Argumentationsfiguren mit Blick auf Universalisierungen finden. Boger (2019) unterscheidet zwischen emanzipatorischen und imperialen Bewegungen der Universalisierung. Bei erstgenannten wird aus einem emanzipatorischen Interesse heraus versucht, das Universale zu (er-)finden, um *allen* gerecht zu werden. Dazu werden in Bewegungen der Dezentrierung der hegemonialen Perspektive falsche Universalismen als solche aufgedeckt. Aus diesem Verständnis heraus könnte man die Bereinigungsversuche also als Versuche lesen, die partikulare Perspektive der Dominanzkultur, die fälschlicherweise für die universale, die neutrale und unmarkierte gehalten wird, als solche zu benennen und eben aus den Konstrukten herauszuhalten, sodass diese tatsächlich universal sind. Imperiale Universalisierungen gehen im Kontrast dazu gewaltsam vor: Das Ziel dieser Bewegungen besteht darin, die eigene partikulare Perspektive zu universalisieren, indem man sie allen anderen aufzwingt. In diesem zweiten Fall besteht die Kernmotivation also darin, das Eigene als das Universale darzustellen, um es in der ganzen Welt auszubreiten.

Auffällig ist in dem dargelegten testpsychologischen Programm die Absenz eines der einfachsten Argumente: Wenn das Ziel darin besteht, Tests zu konstruieren, die in den verschiedensten Kulturen und Sprachen valide und fair sind, warum fragt man dann nicht Forscher:innen aus den verschiedensten Kulturen und Sprachen nach ihrer Perspektive? Die im vorherigen Abschnitt dargelegten Versuche basieren auf der stillschweigenden Annahme, dass nur die Heterogenität oder Homogenität der zu testenden Population relevant wäre. Von der Heterogenität oder Homogenität der testenden und der Tests konstruierenden Population – der Forscher:innen und Praktizierenden – wird geschwiegen. Aus der Perspektive einer intersektionalen Diskriminierungskritik ist die Reflexion der Sprechposition der Forschenden jedoch ein Kernelement: Es sind *deren* Vorstellungen von der jeweils eigenen und fremden Kultur, die in die Testkonstruktion einfließen.

Eine zweite Unterscheidung, die für die Erörterung der motivationalen Ebene bedeutsam ist, ist jene zwischen Förderdiagnostik und Selektionsdiagnostik (für Übersichten/Einführungen siehe z. B.: Simon/Simon 2014; Kornmann 2018): Bedarf es der Fairness, um eine Selektionsentscheidung ‚unanfechtbar‘ zu machen? Oder geht es darum, Verfahren zu finden, die der Förderung in allen Fällen gleichsam dienlich sind? Erstgenannte Motivation entspricht der Haltung, die im Eingangszitat von Lorenz karikiert wurde. Die letzte Frage wiederum klingt nicht aus Versehen schief: Warum sollte man sich auf *ein* standardisiertes, kulturfares Verfahren beschränken, wenn man auch viele verschiedene mit kleinerem, aber dafür passgenauem Anwendungsbereich haben könnte? Selbst wenn es möglich wäre, einen standardisierten und zugleich kulturfairen Test zu entwerfen, spricht bei Vorliegen einer Fördermotivation nichts dagegen und einiges dafür, für jeden singulären Fall die Sinnhaftigkeit verschiedenster möglicher Verfahren abzuwägen (und diese Wahl zu verantworten; s. Abschnitt 2.2).

Das Problem der Fairness von Testverfahren erweist sich dadurch als ein durch gesellschaftliche Verhältnisse hervorgebrachtes: es besteht nur insofern und so lange wie es einen gesellschaftlich produzierten Zwang zu Selektionsentscheidungen (bzw. zur Legitimation dieser) gibt und insofern und so lange wie auch eine in förderlicher Absicht geschehende Diagnostik durch die gesellschaftlichen Verhältnisse mit einer Stigmatisierung einhergeht (gemeinhin als Etikettierungs-Ressourcen-Dilemma bezeichnet). Im Bereich der Sonderpädagogik zeigt sich hier ob der Machtposition von Lehrkräften ein besonderer Bedarf nach Professionalisierung mit Blick auf die Interpretationsspielräume (Kottmann/Miller/Zimmer 2018: 34).

Fernerhin ist aus einer intersektionalen Perspektive nicht nur das intersubjektive Geschehen, sondern auch der institutionelle Kontext kritisch in den Blick zu nehmen:

„Die Intelligenzdiagnostik ist darüber hinaus institutionenorientiert in doppelter Hinsicht: Die Vertreter einer bestimmten Institution, nämlich der Sonderschule, führen Intelligenzdiagnostik durch und legitimieren damit – je nach ihrem gemessenen Ergebnis – einen sonderpädagogischem Förderbedarf der Kinder und die Zuweisung zu ihrer Institution. Das bedeutet eine Definitionsmacht unter der Behauptung einer objektiven Diagnostik, entspricht allerdings dem Festhalten an einer Diagnostik, die sich für das Kind häufig als unscharf, eigenschaftsorientiert, subjektiv, benachteiligend, kontraproduktiv und nicht valide herausgestellt hat“ (Kottmann 2006: 128).

Dies verweist bereits auf den Zusammenhang von Motivation und (institutionellem und kulturellem) *Kontext* — dem Begriff, dem der nächste Abschnitt gewidmet ist.

4 Kontextuelle Dimension: Kultur zwischen Verdinglichung und Individualisierung

Zuletzt soll es um jene Frage gehen, die man typischerweise als definitorische Setzung an den Anfang stellen würde: Was ist ‚Kultur‘ – auf der Ebene der Gegenstandskonstruktion sowie auf der Ebene der Reflexion des methodischen/analytischen Zugangs – in der Testpsychologie und in der Intersektionalitätsforschung? Wie im Folgenden entlang der Unterscheidung in kulturelle Kontexte (4.1) und Kultur als Kontext (4.2) deutlich werden wird, kulminieren in dieser überaus komplexen Konstruktion, die sich daher der pointierten Definition (zumal im Rahmen objektivierender Operationalisierungen) entzieht, alle zuvor dargelegten Überle-

gungen, weswegen sie nun zuletzt verhandelt und zu intersektionalen Perspektiven in Beziehung gesetzt wird (4.3).

4.1 Kulturelle Kontexte

In der ersten Lesart wird der Kontext der Testung sowie der Diagnostik insgesamt als ein kulturell kontingenter anerkannt und berücksichtigt. So plädiert z.B. Möbus (1983) dafür, Test-Fairness nicht *in* der Testkonstruktion, sondern in einer fairen Beziehung zwischen Test und einem externen Kriterium zu suchen. Die übergeordnete Frage lautet hierbei: „Sind die auf seiner Basis angestellten Schlussfolgerungen fair gegenüber Individuen, Gruppen oder Institutionen?“ (Guthke 1972: 25 f. zitiert nach Möbus 1983: 177). Mit dieser Frage öffnet sich die Testpsychologie für die kulturell kontingenten/institutionellen Kontexte, in denen sie zum Einsatz kommt, wie auch Wottawa/Amelang (1980) es taten, als sie betonten, dass insbesondere im diagnostischen Kontext das zu erwartende „Ausmaß an faktischer Bewährung bei einem gegebenen Testpunktwert“ (Wottawa/Amelang 1980: 202) ausschlaggebend ist. Die in einer Kultur situierte Praxis wird dadurch zum entscheidenden Gradmesser. Anstelle von Mittelwertdifferenzen soll die Beziehung zwischen einem Kriterium wie z.B. Schulerfolg oder Fördermöglichkeiten und dem Test als Prädiktor für dieses analysiert werden (Möbus 1983: 177). Unfair wäre ein verbaler Intelligenztest also nicht allein durch unterschiedliche Mittelwerte für verschiedene Gruppen, sondern auch im Falle einer „subgruppenspezifischen Fehleinschätzung“ der Kriteriumswerte (Wottawa/Amelang 1980: 202). Dies wäre z. B. der Fall, wenn ein verbaler Intelligenztest als Prädiktor für allgemeinen Schulerfolg ausgewählt wird, da der Test dieses Kriterium aufgrund des Einflusses der Sprache für anderssprachige Populationen systematisch unterschätzt (Wottawa/Amelang 1980: 203). (Un-) Fairness kann sich also auch durch die falsche bzw. richtige Testauswahl in einem spezifischen Anwendungskontext ergeben. Zur Bewertung einer Beziehung zwischen Test und Kriterium als fair bzw. unfair liegen verschiedene Definition vor (für Übersichten s. Darlington 1971; Möbus 1983). Nach Cleary gilt eine Test-Kriteriumsbeziehung beispielsweise als fair, wenn die vorhergesagten Kriteriumswerte für eine Gruppe durch die allgemeine Regressionsgleichung nicht systematisch über- oder unterschätzt werden. Ist dies der Fall, sollten die Regressionsgleichungen der verglichenen Gruppen identisch sein und sich nicht etwa in ihrer Steigung (slope bias) oder ihrem Schnittpunkt mit der Ordinate (intercept bias) unterscheiden (vgl. Cleary 1968: 115 nach Möbus 1983: 178 f.; vgl. auch Darlington 1971: 74; für eine psychometrische Erklärung der Test-Bias vgl. Furr/Bacharach 2014: 316 ff.).

Ebenfalls im Sinne einer Kritik der Dekontextualisierung der Testdiagnostik argumentiert Sternberg, dass (Intelligenz-)Tests keine isolierten, unveränderbaren Eigenschaften abbilden, sondern deren wechselseitige Interaktionen mit kulturellen Faktoren (vgl. Sternberg 2004: 331):

„Work that seeks to study intelligence acontextually may impose an (often Western) investigator’s view of the world on the rest of the world, frequently attempting to show that individuals who are more similar to the investigator are smarter than individuals who are less similar“ (ebd.: 325).

Er konkludiert, dass man weder kulturfreie noch kulturfaire Tests konstruieren könne, sondern lediglich „kulturrelevante Tests“ (ebd.: 336). Auch hier kommt es nun also ganz zentral auf den Kontext an: *Das Testen selbst ist eine kulturelle Praxis* und stellt einen sehr spezifischen

Kontext zum Einsatz von (mental)en Fähigkeiten dar: „[the] problem was not at the level of mental processes but at the level of applying the processes in specific environmental contexts“ (ebd.: 329). So könnte z. B. der in perzeptuellen Tests wie dem Culture Fair Test vorausgesetzte Umgang mit geometrischen Objekten und abstrakten Problemen durchaus durch kulturelle und insbesondere auch durch schulische Lerneffekte in der Anwendung von Operationen auf einem spezifischen Abstraktionsniveau in einem (verschul)ten Papier-und-Bleistift-Kontext verzerrt sein (vgl. Kubinger 2011; Sternberg 2004; Süß 2003).

Die Konstrukte müssen demnach vervielfältigt werden statt eines davon zu universalisieren und vermeintlich ‚kulturfrei‘ machen zu wollen – denn: was als intelligent gilt, ist kulturabhängig (vgl. Sternberg: 2004). Dadurch werden auch die Kontexte vervielfältigt, in denen sich das zu Testende zeigt (Gardner/Hatch 1989: 6): Eine „Intelligenz-faire“ Testung kann nicht ausschließlich über Papier-Bleistift-Tests erfolgen, sondern muss aus verschiedenen Tätigkeitsformen abgeleitet werden (z. B. sprachliche Fähigkeiten durch Geschichten erzählen statt Vokabeltests, räumliche Fähigkeiten durch Zeichnen oder Zusammenbauen von Gegenständen) (vgl. zur Debatte auch Bathje/Feiss 2014).

4.2 Kultur als Kontext(variable)

In der zweiten Lesart geht es um Kultur als Kontext(variable) in Bezug auf das zu Messende/Testende. Den stärksten Kontrast zu den in 3.1. beschriebenen kulturfreien Ansätzen stellt das multikulturelle Assessment dar, bei welchem kulturelle Einflussfaktoren nicht bereinigt werden sollen, sondern eine explizit kulturspezifische Psychodiagnostik angestrebt wird. Kultur wird als Kontextvariable dabei Teil des Konstrukts. Klient:innen sollen über eine kultursensible Interpretation ihrer psychischen Fähigkeiten bzw. Symptome diagnostiziert werden (vgl. z. B. Martines 2008: 33 ff.). Zu kulturspezifischen Ansätzen zählt zumeist die Vermittlung von kulturellem Grundwissen, wie z. B. kulturelle Werte, Glauben, Identitätskonzepte und *health beliefs* sowie eine Sensibilisierung für Stereotype und Rassismen (vgl. z. B. Dana 1993). Darüber hinaus existieren regelrechte Listen an kulturellen Moderatorvariablen, die je nach Minderheit erhoben werden sollen (ebd.). Im schulpсихologischen Kontext wird der Prozess des multikulturellen Assessments von Martines (2008) leitgebend in verschiedene Phasen eingeteilt. In der ersten Phase sollen durch Befragung des:der Schüler:in Informationen über den kulturellen Hintergrund gesammelt werden (ebd.: 36). In einem zweiten Schritt werden diese interpretiert indem zugeordnet wird, welche Symptome auf individuelle vs. auf kulturell geprägte Umweltbedingungen zurückzuführen sind. Die drängende offene Frage bei diesem Zugang lautet jedoch: Woher soll man das (verlässlich) wissen? Kultur wird in diesem Verfahren verdinglicht, als beschreibbares Objekt bzw. als Faktor *außerhalb* des Selbst phantasiert, über das man unabhängig von einem Selbst, *objektiv* Auskunft geben können soll. Die Fehleranfälligkeit dieses Verfahrens liegt auf der Hand. Aufgrund der Verdinglichung des Kulturellen liegt es nahe, dass die bei diesem Vorgehen auftauchenden Fehler auch auf westliche Konstruktionen von Patient:innengruppen und Kulturen, und somit auf Stereotype, die bloß für objektiv gehalten werden, zurückgehen können. Dadurch können kultursensible Ansätze zu einer „Vertiefung kultureller Fremdheit“ führen, weil es Menschen auf bestimmte Eigenschaften festlegt“ (Amirpur 2016: 41). Die Verdinglichung geht mit einer Essentialisierung einher.

4.3 Vergleichende Abwägung aus intersektionaler Perspektive

Aus einer intersektionalen Perspektive erweist sich der in 4.1 beschriebene Weg der *Anerkennung des Test-Kontextes als kulturell* als eher zu einer Rassismuskritik und einer fairen Diagnostik geeignet, als der in 4.2 dargelegte, der Kultur als Kontextvariable im Test zu objektivieren versuchte. Durch die Verdinglichung von Kultur trug der zweitgenannte Versuch zu einer Essentialisierung bei. Auf eine gewisse Weise stellt er das genaue Gegenteil zum Versuch der Konstruktion eines kulturfreien Tests (3.1) dar, insofern das Kulturspezifische diagnostischer Verfahren darin maximiert wird (statt es minimieren oder sogar eliminieren zu wollen). Zwar wurde die Einsicht gewonnen, dass Kultur kein Labor ist, doch soll sie in der Verdinglichung beschrieben werden *wie* ein Labor. Diese unlösbare Aufgabe wird im sog. multikulturellen Assessment auf die diagnostizierenden Praktiker:innen abgewälzt, die dies gewissermaßen ‚intuitiv‘ bewerkstelligen und entscheiden sollen.

Warum genau erweist sich dieser Weg aus intersektionaler Perspektive als weniger hoffnungsvoll als der in 4.1 dargelegte, der das Testen selbst als kulturelle Praxis versteht? In 2.2 wurden zwei verschiedene Stoßrichtungen intersektionaler Perspektiven unterschieden – eine repräsentationskritische und eine, welche die Notwendigkeit der Vervielfältigung von Repräsentationen und Identitäten betont. Beide richten sich gegen eine Essentialisierung von Kultur. Aus der erstgenannten, dekonstruktiven Perspektive gilt es, essentialistische Konstruktionen von Kultur kritisch zu hinterfragen, da diese Möglichkeitsbedingung von Rassifizierungs- und Ethnisierungsprozessen sind. Aus der zweitgenannten Perspektive, die auf die Repräsentation jener zielt, die durch das Ignorieren der Überkreuzungen von Herrschaftsverhältnissen unsichtbar gemacht werden, ist die Essentialisierung einer jeden Kultur kritikwürdig, da sich die entsprechenden Bilder einer Kultur zumeist als Bilder einer männlich dominierten Mittelschichtkultur erweisen. Auf beiden Wegen wird das verdinglichte/objektivierte Konstrukt ‚Kultur‘ also aufgehoben – durch individualistische Auflösung des im repräsentationalen Akt behaupteten Gemeinsamen aller Angehörigen dieser Kultur oder aber durch eine dekonstruktive Erosion, welche die gemessene ‚Kultur‘ als Artefakt (um einen anschlussfähigen statistischen Begriff zu verwenden) entblößt. In einer vielleicht etwas überpointierten Gegenüberstellung ließe sich zusammenfassen: In 3.1 richtete sich das testpsychologische Objektive gegen das Kulturelle (insofern das Kulturelle (inter-)subjektiv und verzerrend ist). In 4.2 sollte das Kulturelle durch Verdinglichung objektiviert werden. In 4.1 hingegen wurde anerkannt, *dass das testpsychologische Objektive selbst eine kulturelle Konstruktion ist*. Testen und Messen sind kulturspezifische Praktiken – zumal in der Papier- und Bleistift-Variante. Mit dieser Wendung nähert sich die Testpsychologie der auch in der Intersektionalitätsforschung etablierten Forderung nach einer Standpunktreflexion an, bei der das Forschen selbst als situierte Praxis in den Blick gerät (Haraway 1995; Collins 2013).

5 Fazit: Intersektionale Perspektiven auf Psychodiagnostik

Das Testen und Messen selbst ist eine kulturspezifische Praxis. Von dieser in 4.1 mit der Testpsychologie hergeleiteten und in 4.3 aus intersektionaler Perspektive als der hoffnungsvollere Weg beschriebenen Kernthese aus lassen sich nun im Fazit intersektionale Perspektiven auf Psychodiagnostik pointieren.

Argumentiert wurde zunächst mit der Nicht-Delegierbarkeit von Verantwortung (2.2): Es ist ethisch geboten, die subjektive Dimension des diagnostischen Prozesses nicht zu leugnen oder zu streichen, da dem Kind dadurch das pädagogische Gegenüber entzogen werden würde, das sich ver-antwort-et. Die kulturelle Praxis des Testens und Messens ist eine von Subjekten getragene Situation. In dieser wird die Motivation hinter dem diagnostischen Prozess fraglos für das Kind spürbar – auch wenn sie unausgesprochen bleibt oder kaschiert wird. Wenn diese Motivation sich im Rahmen einer förderlichen pädagogischen Beziehung entfaltet, wird dies die Auswahl der diagnostischen Verfahren dahingehend beeinflussen, dass sich die Frage nach der *test-immanenten* Fairness immer dann erübrigen kann, wenn nicht strukturelle Zwänge – wie etwa das Etikettierungs-Ressourcen-Dilemma – zu einer selektierenden Entscheidung und/oder einer klaren Kategorisierung nötigen (3.2).

Dabei wird, wie in 4.1 mit Möbus (1983), Wottawa und Amelang (1980) und anderen skizziert, anerkannt, *dass Fairness in der Relation zwischen dem ausgewählten Test und den involvierten Subjekten dieser kulturellen Situation verortet ist*. Diskriminierungskritische Betrachtungen sollten sich folglich dieser Relation widmen (statt sich in Modellierungen zu verrennen, deren Pseudo-Komplexität die Ertraglosigkeit dieser Versuche eher verdeckt als das Problem tatsächlich zu lösen). Testen und Messen in diesem Sinne als kulturelle Praxis zu verstehen, bedeutet auch, das potentiell Fremde, Verstörende und Verängstigende einer solchen Situation nicht aus dem Blick zu verlieren, sondern genau hier mit einer Kritik des dominanzkulturellen Habitus zu beginnen.

Benachteiligungen, die durch die und in der kulturell geprägten Testsituation selbst entstehen, wie z. B. durch den Stereotype Threat (Steele/Aronson 1995), der Angst einem Stereotyp zu entsprechen, sind weitere Belege dafür, dass sich das Problem nicht durch test-immanente Verschlimmbesserungen lösen lässt. Da diese Befürchtungen kognitive Kapazitäten beanspruchen, können sie in schlechteren Leistungen resultieren (Bathje/Feiss 2014).

Schwarz (2013: 14) betont daher, dass Intersektionalität in der Psychologie auf Interdisziplinarität angewiesen ist. (Test-)Psychologie-immanent lässt sich Diskriminierung nicht beenden. Darum weiß auch die Testpsychologie – was mit Blick auf die im Eingangszitat persiflierte Haltung bedeutet: Es wäre eine fragwürdige Ehrung für die Testpsychologie, diese durch die Hoffnung darauf zu überhöhen, dass sie mit neuen oder ‚innovativen‘ Testverfahren Diskriminierungsfreiheit als Facette von Gerechtigkeit herstellen könne. Als ethisches Antwortgeschehen kann sich die Hoffnung auf Gerechtigkeit nur auf das Inter-Subjektive und Relationale beziehen.

Dies bedeutet auch, dass es (zumindest diesbezüglich) keiner weiteren Runde Methodenstreit bedarf: Sowohl die Kritiker:innen dieser quantitativen Verfahren als auch jene, die sie verteidigen, reden am eigentlichen Problem – der Diskriminierung rassifizierter und/oder migrantischer Schüler:innen – vorbei. Um den Diskursraum zu betreten, in dem Testen/Messen als eine kulturelle Praxis betrachtet wird, in der rassistische/intersektionale Diskriminierung sowohl reproduziert als auch unterbrochen werden kann, müssen wir als Subjekte die Verantwortung für diese kulturelle Praxis übernehmen. Erst dann können wir sie ändern.

Eine intersektionale Kritik an der Reproduktion von Diskriminierung in diagnostischen Verfahren/Testverfahren verliert sich daher nicht in einer Kritik einzelner Tests, seien diese nun qualitativ oder quantitativ. Stattdessen hinterfragt eine intersektionale Kritik erstens die Dynamiken der Verantwortungsverschiebung, die mit ermüdender Verlässlichkeit dafür sorgen, dass sich am Ende niemand für die Reproduktion struktureller und institutioneller Diskriminierung verantworten muss. Zweitens hinterfragt sie die Motivationen hinter dem Ein-

setzen von Tests, also ob diese in diskriminierender oder anti-diskriminierender Absicht in diesem Kontext Verwendung finden. Und drittens kann sie auch die Kultur des Testens und Messens als eine hegemonial gewordene Kultur in den Blick nehmen, insofern diese durch ihre Nähe zu neoliberalen Regimen der Subjektivierung *per se* einer kritischen Betrachtung unterworfen werden sollte.

Literatur

- Amirpur, Donja (2016): Migrationsbedingt behindert? Familien im Hilfesystem. Eine intersektionale Perspektive. Bielefeld: transcript.
- Anastasi, Anne (1964): Culture-fair Testing. In: Educational Horizons, 43(1), S. 26–30. <http://www.jstor.org/stable/42923464>
- Arendasy, Martin/Sommer, Markus/Feldhammer-Kahr, Martina/Freudenthaler, H. Harald/Punter, Fritz Joachim/Rieder, Anita (2018): Fairness als zentrale Herausforderung moderner Annahmeverfahren. In: Zeitschrift für Hochschulentwicklung, 13(4), S. 37–55. <https://doi.org/10.3217/zfhe-13-04/03>
- Arvey, Richard D. (1972): Some Comments on Culture Fair Tests. In: Personnel Psychology, 25(3), S. 433–448. <https://doi.org/10.1111/j.1744-6570.1972.tb00828.x>
- Bathje, Geoff J./Feiss, Carol (2014): Culture-Fair Tests. In: Teo, Thomas (Hrsg.): Encyclopedia of Critical Psychology. New York, NY: Springer, S. 358–361. https://doi.org/10.1007/978-1-4614-5583-7_371
- Benetka, Gerhard (1997): „Im Gefolge der Katastrophe...“ Psychologie im Nationalsozialismus. In: Mecheril, Paul/Teo, Thomas (Hrsg.): Psychologie und Rassismus. Reinbek: Rowohlt, S. 42–72.
- Boger, Mai-Anh (2019): Wer partizipiert an wessen Bildung? Einsatzpunkte einer universalismuskritischen Bildungstheorie. In: Zeitschrift für internationale Bildungsforschung und Entwicklungspädagogik, 42(3), S. 4–10. <https://doi.org/10.31244/zep.2019.03.02>
- Boger, Mai-Anh (2022; i.E.): Mit dem Schwersten beginnen – Über Inklusion und Intersektionalität. In: Zeitschrift für inklusion-online.net; Ausgabe zu Inklusion und Intersektionalität.
- Cattell, Raymond B. (1940): A Culture-Free Intelligence Test. I. In: Journal of Educational Psychology, 31(3), S. 161–179. <https://doi.org/10.1037/h0059043>
- Collins, Patricia Hill (2013): On Intellectual Activism. Philadelphia: Temple University Press.
- Dana, Richard Henry (1993): Multicultural Assessment Perspectives for Professional Psychology. Boston: Allyn & Bacon.
- Crenshaw, Kimberlé (1989): Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics. In: University of Chicago Legal Forum, 1(8), S. 139–167.
- Darlington, Richard B. (1971): Another Look at „Cultural Fairness“. In: Journal of Educational Measurement, 8(2), S. 71–82.
- Dutton, Edward/Becker, David/Osman, Habab A./Bakhiet, Salaheldin F./Essa, Yossry A. S./Ali, Hamada A. A./Alqafari, Shehana M. A./Hamdi, Ali H. M. N./Alfaleh, Asma S. H. (2018): The Raven's Test Performance of South Sudanese Samples: A Validation of Criticisms of the Utility of Raven's among Sub-Saharan Africans. In: Personality and Individual Differences, 128, S. 122–126. <https://doi.org/10.1016/j.paid.2018.02.018>
- Furr, R. Michael/Bacharach, Verne R. (2014): Psychometrics. An Introduction. 2. Aufl. Thousand Oaks, California: Sage.
- Gardner, Howard/Hatch, Thomas (1989): Multiple Intelligences Go to School: Educational Implications of the Theory of Multiple Intelligences. In: Educational Researcher, 18(8), S. 4–10. <https://doi.org/10.2307/1176460>

- Gomolla, Mechtild/Radtke, Frank-Olaf (2009): *Institutionelle Diskriminierung. Die Herstellung ethnischer Differenz in der Schule*. 3. Aufl. Wiesbaden: VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-91577-7_2
- Haraway, Donna (1995): Ein Manifest für Cyborgs. In: Hammer, Carmen/Stieß, Immanuel (Hrsg.): *Die Neuerfindung der Natur. Primaten, Cyborgs und Frauen*. Frankfurt am Main: Campus, S. 33–72.
- hooks, bell (1984): *Feminist Theory – From Margin to Center*. Boston: South End Press.
- Kornmann, Reimer (2018): Von der Auslesediagnostik zur Förderdiagnostik: Entwicklungen, Konzepte, Probleme. In: Müller, Frank J. (Hrsg.): *Blick zurück nach vorn – WegbereiterInnen der Inklusion*. Band 1. Originalausgabe. Gießen: Psychosozial-Verlag, S. 207–226. <https://doi.org/10.25656/01:15820>
- Kottmann, Brigitte (2006). *Selektion in die Sonderschule. Das Verfahren zur Feststellung von sonderpädagogischem Förderbedarf als Gegenstand empirischer Forschung*. Bad Heilbrunn: Julius Klinkhardt.
- Kottmann, Brigitte/Miller, Susanne/Zimmer, Marianne (2018): Macht Diagnostik Selektion? In: *Zeitschrift für Grundschulforschung*, 11, S. 23–38. <https://doi.org/10.1007/s42278-018-0008-2>
- Kubinger, Klaus D. (2011): Testorientierte Leistungsdiagnostik: Grundlagen und Probleme, Verfahrensinventar und besondere Einsatzgebiete. In: Hornke, Lutz/Amelang, Manfred/Kersting, Martin (Hrsg.): *Leistungs-, Intelligenz- und Verhaltensdiagnostik*. Göttingen: Hogrefe, S. 1–51.
- Lorenz, Alfred L. (1990): „Da halte ich mich lieber raus“ – oder Psychodiagnostik als Berufsaufgabe. In: *Psychologie und Gesellschaftskritik*, 14(2/3), S. 135–151. <https://nbn-resolving.org/urn:nbn:de:0168-ssaar-266140>
- Martines, Danielle (2008): *Multicultural School Psychology Competencies: A Practical Guide*. Thousand Oaks. California: Sage.
- Möbus, Claus (1983): Die praktische Bedeutung der Testfairness als zusätzliches Kriterium zu Reliabilität und Validität. In: Horn, Ralf/Ingenkamp, Karlheinz/Jäger, Reinhold S. (Hrsg.): *Tests und Trends. Jahrbuch der Pädagogischen Diagnostik*, 3. Weinheim: Beltz, S. 155–203. <https://doi.org/10.13140/RG.2.1.4458.9926>
- Petersen, Jennifer (2013): Gender Differences in Identification of Gifted Youth and in Gifted Program Participation: A Meta-Analysis. In: *Contemporary Educational Psychology*, 38(4), S. 342–348.
- Purdie-Greenaway, Valerie/Eibach, Richard P. (2008): Intersectional Invisibility – The Distinctive Advantages and Disadvantages of Multiple Subordinate-Group Identities. In: *Sex Roles*, 59, S. 377–391. <https://doi.org/10.1007/s11199-008-9424-4>
- Raven, John C. (1941): Standardization of Progressive Matrices, 1938. In: *British Journal of Medical Psychology*, 19, S. 137–150. <https://doi.org/10.1111/j.2044-8341.1941.tb00316.x>
- Röder, Mandy/Guderjahn, Lena/Fingerle, Michael/Büttner, Gerhard (2021): Diagnostik im Kontext von Fluchterfahrungen: Eine kritische Betrachtung des dynamischen Testens als Möglichkeit zur Verbesserung von Bildungsteilhabechancen auf Basis eines Rapid Reviews der Literatur. In: *Empirische Sonderpädagogik*, 13(1), S. 21–33. <https://doi.org/10.25656/01:23569>
- Rosselli, Mónica/Ardila, Alfredo (2003): The Impact of Culture and Education on Non-Verbal Neuropsychological Measurements: A Critical Review. In: *Brain and Cognition*, 52(3), S. 326–333. [https://doi.org/10.1016/s0278-2626\(03\)00170-2](https://doi.org/10.1016/s0278-2626(03)00170-2)
- Santos-Hövenner, Claudia/Kuntz, Benjamin/Frank, Laura/Koschollek, Carmen/Ellert, Ute/Hölling, Heike/Thamm, Roma/Schienkiewitz, Anja/Lampert, Thomas (2019): Zur gesundheitlichen Lage von Kindern und Jugendlichen mit Migrationshintergrund in Deutschland. In: *Bundesgesundheitsblatt*, 62, S. 1253–1262. <https://doi.org/10.1007/s00103-019-03012-x>
- Schlack, Robert/Hölling, Heike/Kurth, Bärbel-Maria/Huss, Michael (2007): Die Prävalenz der Aufmerksamkeitsdefizit-/Hyperaktivitätsstörung (ADHS) bei Kindern und Jugendlichen in Deutschland. In: *Bundesgesundheitsblatt*, 50, S. 827–835. <https://doi.org/10.1007/s00103-007-0246-2>

- Schmitt, Manfred/Maes, Jürgen/Seiler, Ulrich (2001): Meßäquivalenz und strukturelle Invarianz von Indikatoren der seelischen Gesundheit bei Ost- und Westdeutschen. In: Zeitschrift für Differentielle und Diagnostische Psychologie, 22(2), S. 87–99.
- Schwarz, Silke (2013): Intersectionality Traveling into Psychology. Portal Intersektionalität. <http://portal-intersektionalitaet.de/uploads/media/Schlüsseltext-Schwarz-FINAL.pdf> [Zugriff am 27.06.2022]
- Simon, Jaqueline/Simon, Toni (2014): Inklusive Diagnostik – Wesenszüge und Abgrenzung von traditionellen „Grundkonzepten“ diagnostischer Praxis. Eine Diskussionsgrundlage. In: Zeitschrift für Inklusion, (4). <https://www.inklusion-online.net/index.php/inklusion-online/article/view/194> [Zugriff am 29.06.2023]
- Steele, Claude M./Aronson, Joshua (1995): Stereotype Threat and the Intellectual Test Performance of African Americans. In: Journal of Personality and Social Psychology, 69(5), S. 797–811. <https://doi.org/10.1037/0022-3514.69.5.797>
- Sternberg, Robert J. (2004): Culture and Intelligence. In: American Psychologist, 59(5), S. 325–338. <https://doi.org/10.1037/0003-066x.59.5.325>
- Süß, Heinz-Martin (2003): Culture Fair. In: Kubinger, Klaus D./Jäger, Reinhold S. (Hrsg.): Schlüsselbegriffe der Psychologischen Diagnostik. Weinheim: Beltz, S. 82–86.
- Testkuratorium (1986): Mitteilung. In: Diagnostica, 32(4), S. 358–360.
- Walgenbach, Katharina (2012): Gender als interdependente Kategorie. In: Walgenbach, Katharina/Dietze, Gabriele/Hornscheidt, Antje/Palm, Kerstin (Hrsg.): Gender als interdependente Kategorie. Neue Perspektiven auf Intersektionalität, Diversität und Heterogenität. 2. durchgesehene Aufl. Opladen/Berlin/Toronto: Barbara Budrich. S. 2–64.
- Wechsler, David (2003): The WISC-IV Technical and Interpretive Manual. San Antonio: TX: Psychological Cooperation.
- Wottawa, Heinrich/Amelang, Manfred (1980): Einige Probleme der „Testfairness“ und ihre Implikationen für Hochschulzulassungsverfahren. In: Diagnostica, 26(3), S. 199–221.

Kontakt:

Maré Kondruß M. Sc.
Universität Regensburg
mare.kondruss@ur.de

Dr. Mai-Anh Boger
Universität Regensburg
mai-anh.Boger@ur.de