

(Aus dem Psychologischen Laboratorium der Freien Universität und Valerius-klinik zu Amsterdam [Vorstand: Prof. Dr. L. van der Horst].)

Zu den Kleinkindertests von Bühler und Hetzer.

Von
Dr. Paul Feitscher.

(Eingegangen am 12. Januar 1939.)

Wenn wir uns fragen, was die Entwicklungstests der Wiener Schule an Neuerungen gebracht haben, so geben uns mehrere Schriften der Verfasser hierüber die erste Auskunft. *Charlotte Bühler, Hildegard Hetzer* und ihre Mitarbeiter haben des öfteren ihre Absichten und die neuartige Fundierung ihrer Entwicklungstests ausführlich dargelegt.

„Die Wiener Entwicklungstests,“ sagt *Ch. Bühler*⁵ (S. 238), „weichen in verschiedener Hinsicht von den sonst gebrauchten Tests ab. Erstens besteht der Großteil dieser Tests in *Spielen*, die das Kind und den Jugendlichen freuen und interessieren. Zweitens wird bei dem Großteil der Tests keine bestimmte Aufgabe gestellt, sondern das Kind sich selbst mit einem Spielmaterial überlassen, und nicht eine in Examenssituation hervorgebrachte optimale Leistung, sondern das natürliche Spiel- und Arbeitsverhalten des Kindes ist das, was beurteilt wird. Hierbei wurden die Spiele und Arbeitsmöglichkeiten so zusammengestellt, daß die allgemeinsten und wesentlichsten Bedürfnisse und Interessen einer bestimmten Entwicklungsstufe dabei zur Auswirkung gelangen...“

Begnügen wir uns vorläufig mit diesem kurzen Zitat und illustrieren die Sache durch einige Testbeispiele:

„Testreihe für das erste Vierteljahr des zweiten Lebensjahres:

Test 1. Stöcke aneinanderreiben, klopfen und lauschen. Das Kind bekommt die beiden Rohrstöcke (zwei kurze Bambusstöcke) zum Spiel, und man beobachtet es während der Dauer von 5 Min., wie es sich mit diesen Stöcken beschäftigt. — Positive Lösung: Das Kind reibt die Stöcke aneinander oder klopft mit ihnen aufeinander.

Test 4. Organisiertes Spiel mit dem Ball. Man setzt das Kind in eine Ecke des Bettchens und rollt ihm den Ball mehrmals zu und wieder zurück. — Positive L.: Das Kind rollt den Ball dem VI. zurück. Das Spiel soll sich einige Male wiederholen.

Test 9. Hohlwürfel ein- und ausräumen. Man gibt dem Kinde die sechs ineinandersteckten Würfel und läßt sie so vor ihm liegen. Das Spiel des Kindes mit diesen Würfeln wird 5 Min. lange beobachtet. — Positive L.: Das Kind nimmt wenigstens einen Würfel heraus, oder es läßt ihn herausfallen, um ihn dann wieder hineinzustecken. Das Herausnehmen und Ineinanderstecken der Würfel soll geschehen, ohne daß man es dem Kinde vormacht.

Testreihe für das fünfte Lebensjahr:

Test 2. Wetteifer im gemeinsamen Spiel. Nach dem Spielregelspiel mit dem Mosaikkasten (Test 1 dieser Reihe) werden die Würfel wieder herausgenommen und

verteilt. VI. schlägt vor, daß nun jeder eine Reihe einsetzen soll und sagt: „Wer wird schneller fertig, ich oder du?“ Dasselbe mit einer zweiten und dritten Reihe, wobei VI. es so einrichtet, daß einmal er und einmal das Kind früher fertig ist. Positive L.: Das Kind geht auf die Situation ein; Wetteifer drückt sich in seinem Verhalten durch beschleunigte Bewegungen, Triumph bei Gewinn, zögernde Zugabe der Niederlage usw. aus.

Test 10. Hampelmann zusammensetzen. Die 6 Teile des (aus einem dünnen Brett ausgesägten und bemalten) Hampelmannes werden dem Kind ungeordnet übergeben und zu freier Verfügung 5 Min. lang überlassen. — Positive L.: Das Kind setzt den Hampelmann zusammen. Bei Armen und Beinen darf rechts und links vertauscht werden¹.“

Diese wenigen Beispiele aus dem 190 Einzeltests umfassenden System für die ersten 8 Jahre können zeigen, daß die Verfasser in Hinsicht auf die Natürlichkeit und Lebensnähe ihrer Tests einen großen Schritt voraus gemacht haben. Die verschiedenen Bearbeitungen der *Binet-Simon*-Staffel und amerikanische und deutsche Testfolgen sind in diesem Punkte lange nicht so fortschrittlich gewesen und leiden an einer mehr oder weniger großen Eingeengtheit, Einseitigkeit und Gezwungenheit der Aufgabenstellung sowie an der Vernachlässigung von Tests, die ein freies, unbefangenes, wenig situationsbedingtes Produzieren beobachten lassen.

Eine zweite Errungenschaft der Wiener Entwicklungstests besteht hierin, daß sie wesentlich auf kinderpsychologischen Einsichten gründen. Sie fanden ihre Fundierung in den wichtigen kinderpsychologischen Arbeiten und ausgebreiteten Studien der Wiener Schule, die sich vor allem auf die typischen Eigenheiten und Entwicklungsfortschritte einzelner Altersabschnitte der Kindheit richteten. So können nun mit diesem kinderpsychologisch unterbauten Testsystem die Kinder sozusagen nach inneren Gesichtspunkten getestet werden, während früher mehr von außen herangebrachte Bestimmstücke entschieden. Bei *Binet-Simon* wurden periphere Leistungssymptome in die Waagschale geworfen. Im *Bühler-Hetzer*-System aber wird getrachtet, „auf Grund genauer Kenntnis des kindlichen Aktionssystems“ wichtige Indizien des psychischen Entwicklungsstandes heranzuziehen. Es ist auf diese Weise einigermaßen gelungen, wesentliche Produktionsdispositionen bestimmter Altersphasen als Kriterium zu gebrauchen.

Neben den zwei genannten Eigentümlichkeiten der Wiener Tests: der Angleichung der Testaufträge an natürliche Spiel- und Schaffenssituationen des Kindes und der Gestaltung der Tests nach wichtigen Tatsachen des kindlichen Entwicklungsverlaufes, sollen hier wenigstens noch zwei andere Neuerungen angeführt werden. Die eine finden wir in dem Prinzip, daß das Prüfungsverfahren nicht nur die Intelligenz des Kindes, sondern „alle Grundrichtungen des menschlichen Verhaltens“ sichtbar machen soll. Diese Absicht der Verfasser, nicht allein die Intelligenz, sondern weit *mehr* zu prüfen, scheint mir ihren guten Sinn zu haben. Erstens natürlich, weil man dann zu einer reichereren und umfassenderen Diagnose

über die Persönlichkeit kommen kann. Zweitens aber dürfte meines Erachtens jedes andere Verfahren, das sich also von vorneherein *einzelnen* Dispositionen, Dimensionen oder Funktionen zuwendet, psychologisch inadäquat arbeiten. Denn bei dem ursprungsnäheren psychischen Leben des Kindes rächt es sich noch stärker als beim Erwachsenen, wenn man den Integrationszusammenhang aller Kräfte und Regungen aus dem Auge verliert. Ein einzelnes Gebiet wie z. B. das Denken läßt sich hier schwerlich als selbständige Größe und als losgelöster Untersuchungsgegenstand betrachten und noch weniger beurteilen. Es muß vielmehr, da es fortwährend vom ganzen Tun und Befinden des Kindes her beeinflußt wird, in ständigem Hinblick auf das ganze psychische Leben erkundet werden. Wenn also die Verfasser darauf ausgehen, die ganze Persönlichkeit mit Inbegriff der geistigen Fähigkeiten zu untersuchen, und ein „Entwicklungsalter“, kein „Intelligenzalter“ bestimmen wollen, dann garantiert uns das mehr psychologische Richtigkeit als eine eindimensionale Untersuchung.

Dem Ausdruck, daß das Testverfahren „alle Grundrichtungen des menschlichen Verhaltens“ prüft, kann ich mich freilich nicht anschließen. Man könnte eventuell sagen „gewisse Hauptrichtungen“. Denn wem ist es möglich, alle Grundrichtungen menschlichen Verhaltens zu nennen, ohne daß er viel zu enge und eigenmächtige Überlegungen anstelle? Und wenn wir eine gültige Tafel der Grundrichtungen hätten, würden wir sie dann alle mit Hilfe von Tests untersuchen können? Wir wollen die Möglichkeiten einer Testuntersuchung auch bei Kindern nicht überschätzen und die Grenzen der Testerfassung nicht zu weit stecken. Manche wichtige und tiefe Richtungen des Sozialen, des Begehrrens, der Neigungen usf. können naturgemäß in der Untersuchung nicht wesentlich aufscheinen, sondern müssen nach lebensgeschichtlichen Bewährungen beurteilt werden (Anamnese, eingehender Kontakt). Und schließlich: was für Richtungen, für Dimensionen nennen die Verfasser? Zur Aufstellung eines „Rossolimo-Profiles“ ihrer Testergebnisse gebrauchen sie die Dimensionen: Sinnliche Rezeption, Körperbewegung, Soziales, Umgang mit Material, Lernen und geistige Produktion. Das sind gewiß sehr praktische Einteilungspunkte zur angewandten Kinderpsychologie, aber doch nicht die Grundrichtungen menschlichen Verhaltens.

Als vierte Neuerung sei noch genannt, daß bei der Feststellung der Testergebnisse die soeben genannten sechs Dimensionen unterschieden werden, wodurch die Aufstellung eines psychologischen Profiles im Sinne Rossolimos ermöglicht wird. Die Aufteilung in diese Gebiete soll, über das umfassende Bild des Entwicklungszustandes hinaus, eine qualitative Analyse ergeben.

Diese kurze Besprechung einzelner Prinzipien der Testfundierung mußte dem Folgenden vorangeschickt werden; sie wäre aber unvollständig,

wenn sie die Gesamtfundierung der Tests kennzeichnen sollte. Hierzu sei auf die ausführlichen Schriften der Verfasser verwiesen, die schon bei der ersten Publikation der Wiener Tests umsichtige theoretische und methodische Auseinandersetzungen brachten und die ganze Sache so neuartig und so breit aufbauten, daß eine kurze Besprechung nicht allen wichtigen Gesichtspunkten genügen, sondern nur ihren eigenen Zwecken dienen kann.

Es stand für uns vom Erscheinen der Tests an fest, daß hier sehr Wichtiges für die Praxis der Kinderpsychologie geleistet wurde. Unsere Frage war aber: Wieweit entsprechen die Tests den Absichten der Verfasser? Stimmt alles, wenn man sie in der heutigen Form gebraucht, sind sie allein ausbau- und erfahrungsbedürftig, oder sind eingreifende Änderungen zu wünschen?

Wir begannen die Beantwortung dieser kritischen Fragen bei der Untersuchung von normalen Kindern und Patienten der Kinderpoliklinik und gingen dann nach längerer Erfahrung zur Untersuchung einer ersten Schulkklasse und eines Kinderheimes über, die uns zwar kein großes statistisches Material, aber doch viele Anhaltspunkte und qualitative Ergebnisse lieferte *.

Bericht über eine Schuluntersuchung in Amsterdam zur Frage der Bewährung der Bühler-Hetzer-Tests.

Lotte Danzinger, eine Mitarbeiterin *Charlotte Bühlers*, gab ein Jahr nach dem Erscheinen der „Kleinkindertests“ (Entwicklungstests) von *Bühler-Hetzer* ein Büchlein über einen „Schulreifetest“² heraus, in dem die schon bekannte Testreihe für das 6. Lebensjahr ein wenig erweitert und ausführlich besprochen wurde. Die Ergebnisse einer Untersuchung mit dieser Reihe sollen besonders charakteristisch dafür sein, ob der Prüfling schulreif oder noch unfähig zum Schulbesuch ist. Später hat vor allem *H. Hetzer* Studien über den Entwicklungsstand der Schulneulinge angestellt und die Testreihen für das 5.—7. Lebensjahr mit eigenen Zufügungen als Schulreifetest verwendet. Wir gebrauchten bei der Schuluntersuchung die Methode *Hetzers* und haben die 30 Einzeltests für die genannten Jahrgänge dargeboten.

Die Untersuchung fand im Herbst 1937, kurz nach dem Schuleintritt der Prüflinge statt. Wir untersuchten alle Schüler der ersten Klasse einer Privatschule in Amsterdam, in der gewöhnlicher Grundschulunterricht erteilt wird **. Die Prüflinge gehörten größtenteils wohlhabenden Kreisen

* Die Untersuchung wurde von der Nervenärztin Frau *A. J. de Leeuw-Aalbers* angeregt, die sich die Frage vorlegte, ob sich bei holländischen Kindern die Schulreife mit Hilfe der Wiener Tests feststellen lasse. Ihre praktische Mitarbeit sowie die von Fräulein *E. Westerbeek* halfen zur Aufstellung der vorliegenden Ergebnisse und Überlegungen.

** Herrn Schuldirektor *Onstein* sei auch an dieser Stelle gedankt für sein Entgegenkommen und seine verständnisvolle Förderung der Untersuchungen.

an, in denen den Kindern viel Pflege und versorgte Häuslichkeit zuteil wird. Es waren also sog. „Wohlstandskinder“, um mit einem Ausdruck der Wiener Schule zu reden. Bei solchen Kindern liegt der Entwicklungsquotient erfahrungsgemäß 20—40 % höher als 1⁸.

Es wurden je zwei Kinder gleichzeitig in einem Raum getestet, in dem sich außer den Prüflingen nur die zwei Versuchsleiter befanden. Jedes Kind wurde in zwei Sitzungen untersucht. (Das habe ich mir zur Regel gemacht für alle Testuntersuchungen, bei denen kein ungefähres Urteil genügt. Man soll die Probanden in mindestens zwei Testsituationen mitmachen und einschätzen lernen, da sich dann oft bedeutende Variationen und Fingerzeige zur Deutung der Ergebnisse ergeben können, vor allem weil bei der zweiten Untersuchung die Momente der Fremdheit und der Examenssituation besser ausgeschaltet werden können.)

Im großen und ganzen waren die Ergebnisse der Testuntersuchung aufschlußreich und richtig. Es wurden mit einer einzigen Ausnahme — der Prüfling mußte auch zufolge der qualitativen Befunde als „nicht schulreif“ bezeichnet werden — von jedem Kind mehr als 20 von den 30 verwendeten Tests positiv gelöst. Die negativen Teilresultate fielen hauptsächlich auf die Testreihe für das sechste und mehr noch auf die des siebenten Jahres, so daß sich eine einwandfreie, harmonische Aufteilung der Plus- und Minuszeichen ergab. Ferner können wir bestätigen, daß die Tests im allgemeinen ergiebig sind für das Erkennen wichtiger Verhaltensweisen und Charaktereigenschaften und sehr gut zu qualitativen Feststellungen ausgenützt werden können. Das nach einem halben Jahr eingeholte Lehrerurteil — wir sprachen mit der Klassenlehrerin, dem Schulleiter und einer Lehrerin, die die Kinder vom Kindergarten und von Turnspielen her gut kannte — verifizierte unsere Beurteilung und unsere bis dahin verschwiegene, aber aufgezeichnete Prognose.

Einen genauen Entwicklungsquotienten konnten wir nicht erhalten, da wir die neuen zum System gehörigen Tests für das 8. Jahr noch nicht zur Verfügung hatten und drei Testreihen nicht genügen, um das Entwicklungsalter festzustellen. Einzelne Kinder hätten bei vollständiger Prüfung wahrscheinlich ein besseres Resultat erreicht. Vermutlich wäre der Entwicklungsquotient, der jetzt durchschnittlich 1 beträgt (wenn wir von einem auffällig gestörten Kinde absehen), beträchtlich über 1 gestiegen.

Unsere Tabelle der Testergebnisse, in die die Einzelresultate eines jeden Tests bei jedem Kinde eingetragen sind, bietet uns keine Anhaltspunkte zu statistischen Bedenken gegen die Tests. Sie können, soweit sich das bei einer Anzahl von 14 Prüflingen sagen läßt, nicht als zu schwer oder zu leicht bezeichnet werden. Wir können annehmen, daß die Kinder in Amsterdam durch den Schulreifetest *H. Hetzers* ähnlich erfaßt werden und ähnliche Resultate liefern wie die deutschen Kinder. Da uns bei

der quantitativen Auswertung der Tabelle nichts Besonderes auffiel und ich auch aus unten zu nennenden Gründen von einer genaueren quantitativen Analyse grundsätzlich absehen will, sei die Tabelle nur in verkürzter Form wiedergegeben.

Das jüngste der 14 Kinder war 6 Jahre alt, das älteste 7 Jahre und 2 Monate, im Durchschnitt betrug ihr Alter $6\frac{1}{2}$ Jahre. Im folgenden sind die Zahlen der *negativen* Testresultate pro Test angegeben. Im Bestfall sind es 0, im ungünstigsten Fall wären es 14 für jeden Einzeltest.

Reihe für das fünfte Jahr.

1. Unterordnung unter Spielregel	0
2. Wetteifer beim Spiel	1 ?
3. Drei Aufträge ausführen	3
4. Vier von fünf Dingen finden	2
5. 12 Silben und 4 Ziffern nachsprechen .	2
6. Haus, Baum, Tisch nachzeichnen . . .	1 ?
7. Zeichnung benennen	1 ?
8. Geduldspiel	2
9. Hölzchen aus dem Klotz stoßen . . .	1
10. Hampelmann zusammensetzen	<u>1</u>
	14 (?) (10%).

Reihe für das sechste Jahr:

1. (Wettrennspiel)	(vermutlich 3)
2. Randverzierung zeichnen	3
3. Vers mit 16 Silben nachsprechen . . .	5
4. Nachbauen eines Bauwerkes	2
5. Abbildendes Merkmal zeichnen	3
6. Geduldspiel	3
7. Herablassen einer Schnur	1
8. Folge von Bildern verstehen	4
9. Erkennen von Verkehrtheiten	3
10. (Laden und Ware zuordnen)	<u>(3)</u>
	24 (30) (21%).

Reihe für das siebente Jahr:

1. Gewichte ordnen	1
2. Urteilssuggestion	3
3. (Pfennigeinstreichen)	$(\frac{19}{6})$
4. Bildgeschichte nacherzählen	4
5. (Teppichmuster)	$(\frac{19}{6})$
6. (Laube zeichnen)	$(\frac{19}{6})$
7. Bonbon aus Röhre angeln	3
8. (Elektron erklären)	$(\frac{19}{6})$
9. Vergleich in der Vorstellung	3
10. Bildgeschichte mit 4 Bildern	<u>5</u>
	19 (etwa 32) (23%).

Die (eingeklammerten) Tests dieser Serie nach *H. Hetzer*⁴, haben wir nicht verwendet, sondern vorläufig durch andere Tests aus der *Binet-Simon*-Staffel und

einen Produktionstest (Darstellung selbstgewählter Gegenstände mit 6 Legehölzchen) ersetzt. Die ursprünglichen Aufgaben schienen uns aus verschiedenen Gründen nicht geeignet für unsere Untersuchung. Das Wettrennspiel verlangt eine umständliche Auslegung und Lehrzeit, vor allem aber besonders gutes Eingehen und Kontakt von Seiten des Kindes, wenn es bei ihm nicht das Gefühl des Unbefriedigtseins oder Versagens hinterlassen soll. Wenn man diesem Test nicht viel Zeit und individuelle Darbietung widmen kann, dann ist die Gefahr groß, daß eine unangenehme Testsituation entsteht. Darum haben wir ihn ersetzt. Möglich, daß er unter Umständen (in Internaten, bei guter Bekanntschaft mit dem Kind usw.) oder bei Vereinfachung der Aufgabestellung wichtige Ergebnisse liefert. Ähnlich steht es mit dem „Pfennig einstreichen“. Diese zwei Tests haben auch den Nachteil, daß bei ihnen das Verbale eine große Rolle spielt, und zwar sowohl die verbalen Fähigkeiten des Prüflings überhaupt als auch vor allem der augenblickliche sprachliche Kontakt mit dem Versuchsleiter. Die Möglichkeit, daß die Kinder hierbei aus Gründen versagen, die für das Ergebnis nicht maßgeblich sein sollten, scheint mir bedenklich groß. Für den Test „Laden und Ware zu ordnen“ hatten wir keine für Holland geeigneten Bilder, er mußte aus technischen Gründen ersetzt werden. Die Aufgabe „das Elektron“ zu beobachten und zu erklären, schien mir wegen ihrer Unbestimmtheit und weil sie Ratlosigkeit wecken kann, wenig geeignet zu sein; doch ist mein Urteil hierüber nicht abgeschlossen.

Die Aufgabe „Teppichmuster zeichnen“ ersetzen wir, weil sie den Kindern wegen der wenig anreizenden Aufgabestellung und wegen des auch bei positiver Lösung mangelhaften Resultates keine Befriedigung gewährt. Außerdem findet sich ein ganz ähnlicher Test in der Reihe für das 6. Jahr, den wir den Kindern zumindest dann aufgeben müßten, wenn sie beim Teppichmuster versagen. Wir haben aber den Test „Randverzierung zeichnen“ für das 6. Jahr verändert und bereichert durch die Aufgabestellung, nicht drei, sondern fünf Figuren in ständiger Aufeinanderfolge zu zeichnen. Die Ergebnisse hiervon waren zufriedenstellend; die neue Anordnung ist aufschlußreicher und scheint kaum mehr Schwierigkeiten zu bieten als die bekannte — was allerdings noch durch eine eigene Untersuchung genau auszumachen ist.

Auch den Test „Laube nachzeichnen“ ließen wir wegen der sehr begreiflichen Unlust und Verzagtheit der Prüflinge weg. Dieser gewiß interessante Test, der nöherer Studien wert ist, dürfte sich nur unter besonders günstigen Untersuchungs-umständen empfehlen.

Mit dieser Bemängelung einzelner Tests, die wir schon nach kurzen Erfahrungen ersetzen zu müssen glaubten, begaben wir uns schon ins Gebiet der negativen Kritik, das gegenüber der Herausstreichung der Vorzüge des Testsystems gewiß nicht vernachlässigt werden darf. Denn auch in den letzten Arbeiten der Verfasser, in denen ihr Testsystem behandelt wird, etwa in *H. Hetzers „Psychologische Untersuchung der Konstitution des Kindes“*, bleibt trotz des wertvollen Ausbaues der Interpretationsmethoden der Mangel bestehen, daß den zahlenmäßigen Resultaten zu viel Wert beigemessen wird und die Reichweite der Tests über-schätzt wird*.

* Die beiden Verfasserinnen arbeiten seit längerer Zeit unabhängig voneinander weiter am Ausbau der Tests. Von *H. Hetzer* ist nun eine Reihe für das 8. Jahr erschienen. Auch von Frau *Bühler* sind Reihen für die folgenden Jahre ausgearbeitet worden. Ein Vergleich wurde von uns noch nicht durchgeführt.

Was haben wir diesbezüglich gegen das System einzuwenden? Unsere quantitativen Ergebnisse stimmen, wie gesagt, ausgezeichnet mit den Anforderungen, denen sie entsprechen sollen, und mit den Ergebnissen an anderen Wohlstandskindern überein. Hinzugefügt sei, daß auch unsere Untersuchung in einem Tagesheim für Kinder* an 10 Drei- bis Sechsjährigen zahlenmäßig harmonische, psychologisch interpretierbare und, im ganzen gesehen, richtige Ergebnisse lieferte. Ebenso waren wir mit den Resultaten von Untersuchungen an Babys und Kleinkindern zufrieden. Was ist dann noch falsch?

Es könnte so sein, daß mit quantitativ harmonischen Ergebnissen psychologische Unstimmigkeiten einhergehen; daß z. B. gar nicht die wesentlichen psychischen Eigenschaften der Kinder getestet werden, die man mit den Tests zu prüfen wünscht, sondern in Wirklichkeit sekundäre Eigenschaften oder unerwünschte Faktoren entscheiden. Ferner könnte hinsichtlich der verifizierten Richtigkeit der Testergebnisse die Sache so liegen, daß die Resultate für Leistungseigenschaften im groben stimmen, während eine richtige Einschätzung der individuellen Persönlichkeit und ihrer Wesenszüge verfehlt wird. Gehört doch die feinere Verifizierung von Psychogrammen zu den besonders schwierigen psychologischen Aufgaben, die man kaum von Fehlerquellen befreien kann. Denken wir nur an die entgegenkommende Phantasie des Vergleichers, wenn er die Übereinstimmung zweier Psychogramme bemessen will. Jedenfalls besagen für uns die vielfältigen Verifizierungen, welche die Verfasser publizierten, und unsere eigene kleine Bestätigungsprobe nur, daß wir zu *halbwegs* richtigen, zur praktischen Verwendbarkeit hinreichenden Diagnosen kommen können. Sie können aber nicht glaubhaft machen, daß die Wiener Tests ohne weiteres eine gründliche und in allen wesentlichen Einzelheiten richtige Persönlichkeitserfassung verbürgen und daß das Verfahren theoretisch und methodisch überall einwandfrei ist. Aus verschiedenen Gründen müssen wir gegen das Wiener System in seiner heutigen Form noch Bedenken haben.

Sie richten sich erstens gegen einzelne Tests aus dem ganzen System, die unseres Erachtens abgeändert oder ersetzt werden sollten. Einige solcher Tests wurden schon genannt und charakterisiert als Aufgaben, die leicht Befangenheit oder Unlust hervorrufen. Einige andere Tests müssen wir aus anderen Gründen ablehnen, z. B. weil sie zu umständlich sind oder zu wenig Symptomwerte ergeben. Doch sollen diese Einsprüche gegen einzelne Tests hier nicht weiter dargelegt werden, sondern einer späteren Besprechung vorbehalten bleiben, weil sie relativ nicht so eingreifend erscheinen wie die Bemängelung anderer Faktoren.

Die Kritik muß sich nämlich auch noch gegen einige mehr umfassende Konstitutive des Systems wenden. Sie seien punktweise behandelt:

* Auch der Direktion des Heimes am Wingerdweg sei Dank gesagt.

I. Die Testergebnisse sollen nach den Verfassern primär symptomatisch für die Entwicklung sein. Die Leistungen, die das Kind zu vollbringen hat, sollen für eine bestimmte Entwicklungsstufe einen repräsentativen Wert haben. „Durch rechnerische Bewertung des Testergebnisses ist es möglich, das *Entwicklungsalter* eines Kindes zu ermitteln, das angibt, welchem Lebensalter seine Leistungen entsprechen, und im *Entwicklungsquotienten*, der das Verhältnis von Lebensalter und Entwicklungsalter ausdrückt, einen Maßstab für die Beurteilung der Entwicklungsnormalität, Entwicklungsvorsprünge oder Entwicklungsrückständigkeit zu finden“⁴ (S. 3). An einer anderen Stelle, an der die Möglichkeit, das Testverfahren elastisch anzuwenden und zu interpretieren, erörtert wird, heißt es dann: „Solange mit der Testprüfung nichts weiter beabsichtigt wird als die Feststellung des Entwicklungsstandes des Kindes, kann diese mehr oder weniger schematisch durchgeführt werden“ (S. 26).

Unser Einwand lautet: Der auf Grund der Testuntersuchung festgestellte Entwicklungsquotient ist falsch. Die Absicht der Verfasser, mit dem Rüstzeug der Wiener Tests den Entwicklungsstand in einer einzigen Zahl auszudrücken, wurde nicht verwirklicht. *Die erhaltenen Zahlen entsprechen nicht einem psychologisch einwandfrei bestimmten Entwicklungsstand, sondern sind in unzulässigem Maße imaginär.*

Denn viele Verhaltensweisen, die mit den Wiener Tests begutachtet werden, sind nicht so entscheidend repräsentativ für einen bestimmten Entwicklungsstand, wie angenommen wird. Es ist verfehlt, alle ermittelten Testresultate als entwicklungscharakteristisch hinzustellen. Es gibt zwar zahlreiche Tests im System, die wie die Aufgaben mit dem Testmaterial der Hohlwürfel und Bauklötze, entwicklungstypische Fähigkeiten erkennen lassen. Man kann mit diesen Tests wirklich primär Verhaltensweisen eruieren, die sich von dem einen Reifestadium zum nächsthöheren Stadium typisch weiterentwickeln, dazu ungemein charakteristisch sind für bestimmte Altersphasen und die auch nie mehr spurlos abklingen können, sondern die notwendige Basis zu allen folgenden Stufen bilden.

Andere Tests prüfen aber Fähigkeiten und Verhaltensweisen, die keineswegs in überwiegendem Maße entwicklungs- und reifebedingt sind. Denken wir nur einmal an die Aufgaben im Gebiete von „Umgang mit Material“, besonders aber im Bereich des Sozialen. Es leuchtet wohl ein, daß es eine Überschätzung der Altersbedingtheit ist, wenn man wichtige soziale Verhaltensweisen, die doch in größter Abhängigkeit von typologischen, individuellen und lebensgeschichtlichen Momenten stehen, welche nicht durch Altersphasen bedingt sind, als hauptsächlich alterscharakteristisch und reifeabhängig darstellt. Ich will nicht bestreiten, daß alle untersuchten Verhaltensweisen altersabhängige und phasentypische Merkmale an sich tragen; es soll nur hervorgehoben werden, daß viele angerechnete Verhaltensweisen auch durch altersunabhängige Faktoren

bestimmt und gekennzeichnet werden. Die Vernachlässigung dieser Faktoren macht es unmöglich, die Formel anzuerkennen: so viele Prozente vor oder hinter dem altersgemäßen Entwicklungsstand, mit der uns der sog. Entwicklungsquotient präsentiert wird. Psychologisch bietet sich eben bei den Testuntersuchungen kein Entwicklungsbild in Ursache und Erscheinung, sondern ein nicht kurz zu kennzeichnendes Leistungs- und Verhaltensbild, das nur teilweise dem psychischen Entwicklungsstand entspricht.

Mit der teilweisen Entsprechung verhält es sich weiter folgendermaßen: die Größe der Korrelation zwischen dem Quotienten und dem Entwicklungsstand wechselt in großem Ausmaße mit den einzelnen Prüflingen. Bei dem einen Kind entscheiden hauptsächlich die Entwicklungsziege, bei dem andern spielen dagegen andere Faktoren eine große Rolle. Es gibt also Fälle — sie werden vorzüglich bei den jüngsten Jahrgängen vorkommen — deren Entwicklungsstand durch den EQ. gekennzeichnet wird, während bei anderen dieses Diagnosemittel ganz unzulänglich ist. So kann uns z. B. bei der prinzipiellen Beurteilung des EQ. die Tatsache stutzig machen, daß Wohlstandskinder erwiesenermaßen höhere Zahlenwerte erreichen, welche erstens nicht etwa größerer Begabung dieser Kinder zuzuschreiben sind und zweitens auch nicht mit einem Entwicklungsbild zusammenfallen, das mit psychologischer Auslegung als besonders fortgeschritten und reif zu bezeichnen ist. Das psychologische Bild, das wir bei diesen Prüflingen sehen, gleicht nicht etwa dem eines „gewöhnlichen“, ein wenig älteren Kindes, sondern es ist qualitativ anders und unterscheidet sich von allen typischen Reife-stufenbildern⁸.

Insoferne sich keine eigentlichen Intelligenz- und Reifeunterschiede ergeben, müßte der Entwicklungsquotient bei den Wohlstandskindern genau so hoch sein wie bei anderen Kindern — wenn er seinem Namen entsprechen würde.

Zur Erhärting des Einwandes können Beispiele angeführt werden: *H. Hetzer*⁴, „Beispiel 14. M 404. Brigitte A., 6 Jahre, EQ. 1,10. Hat einen erheblichen Entwicklungsvorsprung. Ihre Leistungen sind auf allen Gebieten gut. Sie gibt sich an das, was sie tut, ganz hin, ist allerdings zu einer langdauernden, selbständigen Arbeit nicht fähig, einfallsarm und reaktiv. Sie ist sprachlich sehr gewandt, und ihre Sprache ist sehr gepflegt. Ihr Auftreten ist sicher und formvollendet. Die Wohlgepfletheit ist ihr deutlich anzumerken und dürfte auch die Hauptursache ihrer Entwicklungsvorsprünglichkeit sein. . . . Die Erfolge bei den Testaufgaben sind größtenteils darauf zurückzuführen, daß z. B. die Möglichkeit gegeben wurde, entsprechende Erfahrungen zu sammeln, daß sie reichlich Anregungen empfangen hat. . . . Die Sicherheit und Gewandtheit im Auftreten, die Selbstverständlichkeit, mit der sie mit den Erwachsenen spielt, zeigen, daß ihr diese Spielsituationen vertraut sind. . . .“ Der sog.

Entwicklungs vorsprung, der im EQ. zur Erscheinung kommt, ist hier durch Merkmale gekennzeichnet, die doch eigentlich nur sehr wenig mit einer positiven und reinen Charakteristik der psychischen Entwicklung zu tun haben; und so hat sich hier der Fehler eingeschlichen, daß dasjenige entwicklungs vorsprüngig genannt wird, was einen hohen EQ. herbeiführt, während wir doch lieber den „EQ.“ weniger absolut nehmen wollen und dafür den Begriff der psychischen Entwicklung psychologisch sauber anzuwenden versuchen.

Aufschlußreicher als die Frage der ‚fröhreifen‘ Wohlstandskinder ist in dieser Hinsicht ein anderes spezielles Untersuchungsergebnis von *H. Hetzer*. Bei dem Versuch, eine körperseelische Entwicklungstypologie der 5—7jährigen festzulegen³, kommt die Verfasserin zu dem Ergebnis, daß die Kinder dieses Alters somatisch und psychisch deutlich zwei verschiedenen Typen zugeteilt werden können: der Kleinkindform oder der Schulkindform. Die Kleinkindform entspricht körperlich und seelisch einem jüngeren, die Schulkindform einem älteren Alterstypus. Es handelt sich also hier ausgesprochen um Entwicklungstypen, die nach *Hetzer* tatsächlich auch sowohl leiblich als auch psychisch in markanten Erscheinungen zum Ausdruck gelangen. Nun wäre es zu erwarten, daß der sog. Entwicklungsquotient die Zugehörigkeit zu den beiden Typen klar demonstrieren müßte. Wenn man sich aber die Fälle ansieht, die von der Verfasserin als Musterbeispiele für ihre Typologie verwendet werden, dann zeigt sich folgendes: Im ganzen werden 6 Beispiele für charakteristische Kleinkind- und Schulkindformen besprochen, die natürlich besonders instruktiv sein sollen. Darunter sind zwei Kinder im Alter von 6 Jahren und 4 Monaten, das eine gehört dem ersten Typus an, das andere dem zweiten. Was sagt der EQ.? In beiden Fällen beträgt er 0,92. Zwei andere Kinder, die verschiedenen Typen zugehören und genau gleich alt sind, haben auch einen praktisch gleichen EQ. von 0,80 und 0,81. Nur bei einem einzigen Paar hat die ‚Schulkindform‘ einen höheren EQ. Das sind die drei Musterbeispiele. Bei dieser Sachlage stimmt etwas nicht: Entweder die Typologie oder der Entwicklungsquotient. Nun scheint mir die typologische Charakterisierung und Interpretation der genannten Fälle in bezug auf die psychischen Merkmale der relativen Kindlichkeit oder Reife richtig und überzeugend zu sein, so daß ich den Fehler beim EQ. suche. Nur mit Hilfe einer tieferblickenden Auslegung der Testergebnisse und des *ganzen* Gehabens des Kindes konnte *Hetzer* zu ihrer Typologie der Reifeunterschiede kommen. Sie sagt selbst: „... Als Gesamtergebnis unserer vergleichenden Betrachtung können wir buchen, daß ein erheblicher Unterschied zwischen den Leistungen der Kleinkindformen und der Schulkindformen besteht und daß dieser Unterschied vorzugsweise (!) ein qualitativer ist.“ Da es sich aber um eine Entwicklungsfrage handelt, müßte dieser erhebliche Unterschied auch quantitativ im EQ. klar zum Ausdruck gelangen. Das Beispiel der Wohl-

standskinder, das sich durch ähnliche vermehren ließe, zeigte umgekehrt, daß Faktoren, die nichts mit Entwicklungsanachronismen zu tun haben, den EQ. von der Norm entfernen.

Was sind aber die praktischen Schlußfolgerungen für uns im Hinblick auf die Tatsache, daß das Entwicklungsalter und der Entwicklungsquotient im Wiener Testsystem ihre Namen nicht zu Recht tragen und daß ihr Begriff künstlich ist? Sollen alle zahlenmäßigen Resultate für hinfällig erklärt werden?

Im praktischen Gebrauch habe ich mir hierüber folgendes zurecht gelegt und angewendet: 1. Verschiedene Tests gehören, wie gesagt, verändert oder ersetzt; hiedurch kann das Gesamtergebnis um ein Geringes richtiger werden. 2. Ohne psychologische Interpretation dürfen wir das zahlenmäßige Ergebnis nur als einen schwer in seiner Bedeutung zu übersehenden eigenartigen „Leistungs-“ und „Anpassungsquotient“ nehmen. 3. Die wichtigsten Ergebnisse kann uns erst die qualitative Interpretation der einzelnen Testresultate und der einzelnen und allgemeinen Verhaltensweisen der Prüflinge bei der Untersuchung liefern. 4. Erst nach dieser psychologischen Auslegung sind wir imstande, die zahlenmäßigen Einzelresultate und eventuell den Gesamtquotienten als Ergebnis zu verwerten; es wird aber immer wieder zahlreiche Fälle geben, deren ziffernmäßige „Testnoten“ nicht ins Gewicht fallen dürfen, weil diese Fälle qualitativ so beschaffen sind, daß sie durch das normierte Testinstrument nicht richtig erfaßt werden können. Dann muß man sich aber auch entschließen, von einer zahlenmäßigen Darstellung abzusehen. 5. Falls man aber überhaupt zu einem psychologisch zulässigen und aufschlußgebenden Gesamtquotienten kommen kann, dann ergeben sich noch verschiedene Möglichkeiten: Es gibt Fälle, bei denen man mit Recht von einem EQ. sprechen kann; z. B. meistens, wenn das Profil der Ergebnisse auf allen wichtigen Gebieten annäherungsweise gerade und waagrecht verläuft, so daß nicht das eine oder andere Einzelgebiet einen größeren oder kleineren Faktor zum Gesamtquotienten beisteuert als die anderen Sondergebiete. Wenn aber Sondergebiete stark aus dem Bilde fallen und einen relativ großen Einfluß auf den Gesamtquotienten haben, so haben wir es wahrscheinlich nicht mit einem „Entwicklungsquotienten“, sondern je nachdem mit einem „Intelligenzquotienten“, „Leistungsquotienten“ usw. zu tun. Das sei aber nur andeutungsweise gesagt, da sich hier gewiß keine rein quantitativen Regeln geben, sondern nur nach vorangegangener Analyse des Falles aus psychologischem Verständnis heraus einige rechnerische Überlegungen anwenden lassen. 6. Dies alles soll keineswegs besagen, daß wir im allgemeinen die quantitativen Feststellungen verwerfen oder vernachlässigen wollen; wir halten sie vielmehr für wichtig und aufschlußreich dort, wo sie in kritisch abgegrenztem Bereich angewendet werden können. Bei unseren Untersuchungen in Klinik, Schule und Kinderheim konnten wir in den meisten Fällen zu

einem umgrenzt zulässigen Quotienten gelangen, der ein wichtiges Diagnostikum bedeutete.

II. Ein anderes Kapitel der kritischen Besprechung des Wiener Testsystems muß sich mit der Einteilung in die verschiedenen Gebiete oder Dimensionen befassen. Der hauptsächliche Einwand ist hier, daß die Zuordnung der einzelnen Tests zu den verschiedenen Gebieten sehr anfechtbar ist. Man stößt in der Literatur über die Wiener Tests oft auf Widersprüche in der Zuordnung der Tests zu einzelnen Gruppen wie folgende: Der Test VI/2, „Randverzierung auf Aufforderung korrekt zeichnen“ wird der Dimension „Soziales Verhalten“ zugewiesen (¹ und ⁴). Die Befremdung, die diese Zuordnung erweckt, wird unseres Erachtens durch folgende Analysen der Verfasser nicht aufgehoben:

„Diese Aufgabe stellt an die Aufgabewilligkeit des Kindes und an seine *Selbständigkeit* bei der Aufgabendurchführung sehr hohe Anforderungen, denn das Kind muß eine lang dauernde und wenig reizvolle Tätigkeit ohne Hilfe durchführen. Wo der *Leistungswille* eines Kindes gering ist, bricht es die Aufgabe ab, sobald es dieser überdrüssig wird. Es fragt: „Darf ich nicht aufhören“, „geht das noch lange“, seufzt und stöhnt, um seinen geringen Willen zur Fortsetzung kundzutun. Für Vorhaltungen, daß sie fertigmachen müßten, sind sie unempfänglich, sie sind aber sofort bei der Sache, wenn ihnen eine Belohnung in Aussicht gestellt wird. Die *Selbständigkeit* zeigt sich in der Unabhängigkeit vom Prüfer bei der Durchführung. Selbständige Kinder arbeiten, ohne sich um den Erwachsenen zu kümmern, zu Ende. Unselbständige Kinder legen das Blatt oft nach jeder Wiederholung der 3 Figuren dem Prüfer vor, sie arbeiten an der Ecke nicht weiter, sondern warten, bis er ihnen ermunternd das Blatt umdreht.“

Spielerisches Verhalten tritt häufig auf, ebenso zeigt sich die *Leichtermüdbarkeit* in einem Nachlassen der Arbeitsintensität und in gehäuften Fehlern bei fortschreitender Arbeit. Konzentration, Ausdauer und Schwankungen der Aufmerksamkeit sind deutlich an Verhalten und Güte der Leistung zu beobachten. Manche unruhige, immer abgelenkte Kinder führen die Aufgabe nur dann zu Ende, wenn der Prüfer sie ständig bei der Sache hält.

Organisation der Arbeit und Arbeitstempo zeigen sich bei dieser Aufgabe darin, wie das Kind sich das Zeichenblatt beim Anlangen an den Ecken zurechtlegt, ob es sehr langsam oder schnell arbeitet. Ebenso ist geringe *Merkfähigkeit* erkennbar, wenn manche Kinder oft bei der dritten und vierten Seite der Randverzierung noch immer die Figuren abzeichnen, sie noch nicht sicher aus dem Gedächtnis hinsetzen. Die Aufgabe stellt an die *Handgeschicklichkeit* erhebliche Anforderungen. Gerade das Ausführen der kleinen Zeichenbewegungen, vor allem die eckige Strichführung des Dreieckes, fällt handgeschickten Kindern besonders schwer, sie hängt auch von dem Stand der körperlichen Entwicklung ab und wird von den Kindern, die noch dem körperlichen Entwicklungstyp nach der Kleinkindform angehören, nur unvollkommen durchgeführt. Sorgfalt und Nachlässigkeit bei der Strichführung sind ebenso zu erkennen wie bestimmte *Ausdruckswerte* der Zeichnung. Die hauchdünnen schwachen Linien der Figuren, die viel kleiner sind als die vorgezeichneten, finden sich zumeist bei schüchternen, sensiblen Kindern, die dicken Striche und besonders großen Figuren bei robusten Kindern. Bei sehr zerfahrenen, unruhigen Kindern gibt eine erhebliche Ungleichmäßigkeit der Figuren und eine Unzahl von ungewollten oder unnötigen Strichen ein deutliches Bild ihres Zustandes“ (Hetzer⁴). Eine andere Analyse des Testes „Randverzierung zeichnen“

lautet: „In erster Linie wird geprüft Ausdauer und Konzentration; außerdem wird geprüft: Handgeschicklichkeit, Gedächtnis für Gestalten“ (*Danzinger* ²).

Wenn wir nun die Probe machen und alle wichtigen Bestimmstücke der vorhergehenden Analysen daraufhin ansehen, was sie mit der Dimension „Soziales Verhalten“ zu schaffen haben, so kommt heraus, daß unmittelbar nur die Aufgabewilligkeit, aber auch sie nicht zur Gänze, dem sozialen Gebiet angehört. Vielleicht sind auch „Selbständigkeit“ und „Leistungswillen“ zum Sozialen in Beziehung zu bringen, aber das liegt nicht ohne weiteres auf der Hand. Und das meiste gehört gänzlich anderen Gebieten an. Derartige Verlegenheitszuordnungen finden sich mehr, und zwar unseres Erachtens in solchem Ausmaße, daß man sie nicht kurzerhand korrigieren kann. Ich wüßte keinen Weg, um alle Tests einigermaßen einwandfrei auf die sechs Gebiete zu verteilen. Die Fehlerhaftigkeit bei der Kennzeichnung der Gebietszugehörigkeit ist analog zu der künstlichen Verallgemeinerung des Entwicklungsbegriffes. Die zur Frage stehenden psychischen Verhältnisse werden hier wie dort nicht mit einer angemessenen Charakteristik versehen, sondern in eine begriffliche Zwangsjacke gesteckt. Die angemessene Charakteristik findet sich wohl oft in der nachträglichen Interpretation, welche die Verfasser, besonders *H. Hetzer*, in späteren Schriften für viele Tatbestände geben. Merkwürdigerweise läßt man aber daneben die zur Gewinnung von zahlenmäßigen Ergebnissen und Kurven bestimmten unrichtigen Kennzeichnungen weiter bestehen, so daß, wie hier gezeigt wurde, aus den Worten der Verfasser selbst Widersprüche herauszulesen sind. Wir dürfen uns hier keineswegs mit der Meinung begnügen, daß, wenn auch Einzelheiten unscharf und schief gekennzeichnet werden, durch wechselseitige Kompensation der Ungenauigkeiten das globale Urteil stimmen wird. Erstens ist es gewiß nicht oft der Fall, daß ein dermaßen zustande gekommenes globales Urteil richtig ist, und zweitens ist in Sachen der psychologischen Charakteristik an keiner, auch nicht an „untergeordneter“ Stelle eine schiefe Kennzeichnung zu dulden.

Es sind aber nicht nur in Hinsicht auf die unangemessene Zuordnung vieler Tests zu den sechs Dimensionen Bedenken zu äußern, sondern auch in bezug auf die statistische Anrechenbarkeit der Ergebnisse auf einzelnen Gebieten. *A. Huth* hat eingewendet, daß für die einzelnen Jahrgänge Testgruppen vorgesehen sind, die einen ungleichen Anteil an den 6 verschiedenen Dimensionen haben. So ist z. B. für $1\frac{1}{2}$ -jährige ein einziger Test zu finden, der zum Gebiet „Umgehen mit Material“ gehört, während für 2jährige vier solcher Tests in das System aufgenommen sind. Das $1\frac{1}{2}$ -jährige Kind schneidet also testmäßig schon gut ab, wenn es alle Materialtests der vorangehenden Altersphasen plus 1 Test seines wirklichen Alters löst, während das 2jährige der schwereren Anforderung entsprechen soll, alle vorangehenden plus 4 altersgemäße Tests

zu beherrschen. Durchwegs ist die Verteilung aller Gruppen nicht regelmäßigt. Das führt natürlich zu statistischen Unzulänglichkeiten. Im Testsystem für die Motorik von *Oseretzky** — das in anderer Beziehung weit größere Mängel aufweist als das *Bühler-Hetzer*-System — ist diese Angelegenheit vorbildlich behandelt: *Oseretzky* unterscheidet für sein Staffellsystem auch 6 verschiedene Dimensionen; er hat indessen für jeden Jahrgang 6 Tests vorgesehen, von denen regelmäßig der erste dem Gebiete A entspricht, der zweite dem Gebiete B, der dritte dem Gebiete C usw., wodurch, was dies betrifft, quantitative Darstellung unter zulässigen Bedingungen möglich wird. Für die Wiener Tests gilt dies nicht, obgleich der letztgenannte Mangel einigermaßen hierdurch ausgeglichen wird, daß jedes Kind mit den Testgruppen von mindestens drei Jahrgängen zu prüfen ist.

Derselbe Umstand, daß nämlich eine Testbatterie von drei bis sechs Jahrgangsgruppen verwendet wird, vermag aber eine andere Lückenhaftigkeit der Gebietzuteilung weit weniger zu kompensieren. Es handelt sich um die Tatsache, daß die einzelnen Gebiete für sich genommen nur lückenhaft und oft nur unzureichend erfaßt werden, anders ausgedrückt, daß der Symptomwert der Tests eines Gebietes für dieses Gebiet oft nur einseitig und beschränkt ist. Betrachten wir z. B. die drei Testreihen, womit 5jährige zu untersuchen sind, in bezug auf die Gruppen „Materialbeherrschung“ und „Soziales“. Die Tests für jene lauten: Bauwerk benennen — Zeichnung benennen — Abbildendes Merkmal zeichnen — Geduldspiel: eine Maus in die Falle — Geduldspiel: zwei Mäuse in die Falle. Das sind die Tests, an denen der Fünfjährige seine Betätigung am Material bewähren muß und die nach einer einfachen Plus-Minusnotierung seinen Entwicklungsstand auf diesem Gebiete erfassen sollen. Wir können sehen, daß die Tests wohl *etwas* untersuchen, was zu diesem Gebiete gehört; aber diese Stichproben scheinen uns — seien sie noch so entwicklungspsychologisch unterlegt — doch nicht wesentlich das Gebiet der Materialbeherrschung zu erfassen. Dasselbe gilt für die Sozialtests dieses Alters: Sprachliche Formulierung von Plänen — 200 Blättchen auf Aufforderung sortieren — Sittliche Bewertung dargestellter Handlungen — drei Aufträge ausführen — Wetteifer im gemeinsamen Spiel — Unterordnung unter Spielregel (zweimal, bei verschiedenen Spielen), Randverzierung auf Aufforderung zeichnen. Ist das soziale Gebiet hiemit in seinen Grundzügen erfaßt? Ich glaube nicht. Ähnliches gilt auch für das Gebiet der Körperbewegungen bei jüngeren Jahrgängen.

Zum Teile scheint eine Eigenartigkeit der Testanlage an diesem Übelstand Schuld zu tragen. Es gibt nämlich, wie die Beispiele zeigen, einige Tests, die zwei- oder dreimal wiederkehren bei verschiedenen Jahrgängen

* Kürzlich kritisch besprochen von *A. J. de Leeuw*: Nederl. Tijdschr. Psychol. 6, 215 (1938).

und dann nur quantitativ etwas schwieriger gemacht wurden. Z. B. das Geduldspiel, das Nachsagen oder das Finden von 2, 3 oder 4 versteckten Dingen. Durch diese quantitativen Mehrleistungsproben wird die Zahl der das ganze Gebiet bestreichenden Tests noch verringert und die Gebietserfassung beschränkt. Allgemein: Man erhält unterschiedlich interessante und wichtige Stichproben, aber nicht das Umfassend-Wesentliche, was uns durch die Gebietsnamen gesagt oder suggeriert wird.

Man kann alledem entgehen, indem man die Resultate auf einzelnen Gebieten nicht für allgemeingültig hält und nicht quantitativ und zeichnerisch profilmäßig darstellt, sondern sich auf eine psychologisch beschreibende Wiedergabe beschränkt, die aber alle sechs Gebiete berühren muß. Der Verzicht auf das Profil, das so großen Brauchwert hat, ist natürlich nicht leicht, scheint aber vorläufig geboten. Wenn man es aber doch verwenden will, dann dürfen die Testresultate nicht schematisch nach der Zuordnungstabelle eingezeichnet werden, sondern sollen erst nach einer dem individuellen Falle entsprechenden psychologischen Auslegung ihres Zustandekommens beurteilt und zugeteilt werden.

Mit dieser kleinen kritischen Betrachtung könnten wir über den zweiten, dem Dimensionsproblem geltenden Einwand im Rahmen dieser Studie das Nötige dargelegt achten. Da aber über diese Frage in jüngster Zeit eine spezifische längere Studie erschienen ist⁶, soll noch mehr dazu gesagt werden. *R. Hofstätter* bewertet in seinem Artikel: „Was besagen Testergebnisse? — Ein Beitrag zum Dimensionsproblem der Entwicklungstests“, die Frage mittels einer statistischen Methode, der Korrelationsmethode und Faktorenanalyse nach *Thurstone*. Diese Methode beruht auf der Tatsache, daß sich die Verwandtschaft der verschiedenen Tests einer Testbatterie in mehr oder weniger festen Kopplungen der positiven oder negativen Einzelresultate ausdrückt. *Hofstätter* untersuchte so die Testresultate von 100 Kindern, die mit der Testreihe für das 6. Jahr geprüft worden waren, und stellte mit Hilfe rechnerischer Vergleichung der Ergebnisse der 10 Einzeltests fest, daß diese untereinander dermaßen korrelativ sind, daß man statistisch vier gemeinsame Faktoren unterscheiden kann. Den ersten Faktor nennt er — jetzt psychologisch interpretierend — „Erfassung von Zusammenhängen“. Er fällt ungefähr mit der Dimension „Geistige Produktion“ zusammen, wobei es sich zeigt, daß 3 Tests dieses Gebietes deutlich dieselbe Fähigkeit prüfen. Es sind dies: Verkehrtheiten eines Bildes erkennen, kausale Folge von Bildern verstehen und sich durch Herablassen einer Schnur eines Dinges bemächtigen. Der vierte Test der Altersreihe, welcher von den Verfassern auch dem Gebiete „Geistige Produktion“ zugeordnet wird („Waren zusammenstellen“), zeigt eine recht geringe positive Korrelation zu den drei andern, während die Tests „Nachbauen“ und „Vers nachsagen“, mit denen laut Theorie keine Verwandtschaft

vorliegen sollte, ziemlich gut mit ihnen korrelieren. Immerhin ist zu sagen, daß die Dimension „Geistige Produktion“ im Sinne von „Zusammenhänge erkennen“, „Beziehungen herausfinden“, deutlich zu Recht unterschieden wurde und daß nur eine kleine Veränderung in der Testanalyse und Testzuteilung nötig ist, um hier eine richtige Gebietsgruppe zu erhalten. Ferner unterscheidet *Hofstätter* die Gruppen: 2. Die Verwendung von Erfahrungen. 3. Der soziale Kontakt. 4. Die körperliche Aktivität. — Diese weiteren drei Dimensionen stimmen mit den theoretischen von *Bühler-Hetzer*, welche mit ihnen der Erwartung nach zusammenfallen sollten, nur sehr wenig überein. Nirgends mehr findet sich durch *Hofstättters* rechnerische Befunde das spekulative System der sechs Dimensionen so weitgehend bestätigt wie bei der „geistigen Produktion“. Jedenfalls ist es so, daß, wenn sich auch noch verschiedene interessante Teilkorrelationen ergeben, die Übereinstimmung des Befundes mit der Erwartung, die durch die theoretische Aufstellung der Verfasser nahegelegt wird, keineswegs befriedigend ist. Es sind gewiß noch einschneidende Korrekturen anzubringen. — *Hofstätter* selbst ist meines Erachtens auch *noch lange nicht zu einer praktisch befriedigenden Lösung gekommen* — was seine Darstellung auch gar nicht besagen will —, obwohl er eine Reihe von wichtigen Auslegungen und Korrekturen gibt, die sehr beachtenswert sind. Die psychologische Interpretation seiner Befunde, die ihn dann zu der Namengebung der vier Gebiete führt, erscheint, so interessant und fein sie auch ist, nicht unbedingt einleuchtend und zutreffend. *Außerdem hat er nur die Ergebnisse, die an einem Jahrgang mit einer einzigen Testreihe gewonnen wurden, verarbeitet* und sagt selbst, gleichfalls einschränkend: „... Man wird sich auch hier angewöhnen müssen, die Ergebnisse der Analyse niemals unabhängig von der verwendeten Untersuchungsgruppe zu formulieren. Es ist durchaus möglich, daß dieselben 11 Tests, wenn man sie einer Gruppe von um 1 Jahr jüngeren oder um 1 Jahr älteren Kinder vorlegt, jedesmal zu einem ganz andersartigen System von Verwandtschaftsbeziehungen und zu einer ganz andersartigen Dimensionsstruktur führen.... Es wäre verfehlt, wollte man auf Grund dieser Erwägungen die Faktorenanalyse als solche ablehnen, vielmehr wird man immer im Auge zu behalten haben, daß die funktionalen Voraussetzungen, auf Grund deren ein und dieselbe Leistung zustande kommt, unter Umständen recht verschiedener Art sein können. Der alte Satz: si duo faciunt idem, non est idem, darf zumindest als Möglichkeit in der Leistungspsychologie nie übersehen werden. — Es ist verständlich, daß ein Verfahren, das den Aufbau von Tests herausarbeitet, zu dem Irrtum verleitet, als wäre damit eine Angabe gewonnen, die von den Menschen, die diese Tests zu lösen hatten, unabhängig ist. Genau betrachtet wäre ein derartiges Resultat recht uninteressant und im Grunde auch unbiologisch. Vielmehr muß immer und immer wieder betont werden, daß die gewonnenen Aussagen sich stets auf eine bestimmte

Gruppe von Menschen beziehen. Aber gerade darin erblickt der Verfasser einen besonderen Wert dieser Arbeitsweise, da so Aussagen über die zwischen bestimmten Gruppen bestehenden Unterschiede möglich werden.“ (Verfasser — *Hofstätter* — „hält es für eine verlockende Aufgabe, den Alterswandel einmal daran zu studieren, wie für ein und dieselbe Leistungsreihe allmählich immer andere Komponenten maßgeblich werden“. Man denke nur an alles, was der Automatisierung zugänglich ist!) Diese Ausführungen lassen ohne weiteres erwarten, daß ebenso wie man beim Vorlegen der Tests an verschieden alte Kinder jedesmal zu einem andersartigen System von Verwandtschaftsbeziehungen und Dimensionsstrukturen kommt, dasselbe der Fall ist, wenn man die Tests ungleichen Typen oder eigenartigen Individuen darbietet. Hiermit glaube ich meine Ansicht gestützt zu sehen, daß erst die psychologische Auslegung am Einzelfall uns eine psychologisch natürliche Beurteilung ergeben kann.

Ferner dürften sich *Hofstätters* Ergebnisse hinsichtlich einer Korrektur des spekulativen Dimensionssystems (die von ihm nur als Nebenergebnis seiner Studie gebucht wird) folgendermaßen zusammenfassen und deuten lassen: 1. „Der Hauptunterschied gegenüber der *Bühlerschen* Theorie besteht darin, daß auf Grund der Faktorenanalyse kein einziger Test nur einer einzigen Dimension zugeordnet werden kann, wie dies in der spekulativen Einteilung der Aufgaben fast durchwegs geschehen ist“ (S. 92). 2. Es besteht mancherlei Übereinstimmung zwischen Erwartung und Befund. In einigen Einzelheiten hat sich deutlich die Berechtigung der Gebietseinteilung und -Zuordnung gezeigt. Aber ebensogut stellte es sich heraus, daß viele Tests schief gekennzeichnet wurden. (Man kann sich den Analogieschluß erlauben, daß dies auch bei anderen Jahrgängen so ist.)

Weiter müssen wir auf die Studie selbst verweisen und empfehlen ihre Einzelergebnisse bei allen Dimensionsuntersuchungen unbedingt zu Rate zu ziehen.

III. Wenn kritisiert wurde, daß die Eignung, die Allgemeingültigkeit und die Eindeutigkeit vieler Tests von den Verfassern anscheinend überschätzt wird, wenn weiter besonders der Entwicklungsbegriff und der Dimensionsbegriff in der Form, wie er zur quantitativen Begutachtung verwendet wird, künstlich genannt und wenn schließlich gezeigt wurde, daß uns erst eine eingehende Interpretation der Testergebnisse und eine umfassende Beschreibung der ganzen Persönlichkeit instandsetzt, die Entwicklungsmerkmale von den individuellen Persönlichkeitszügen zu unterscheiden und alle Testresultate zu werten, so entsteht die Frage: wie interpretieren die Verfasser selbst die Ergebnisse, wenn sie mit Hilfe der Tests zu ausführlichen Psychogrammen gelangen wollen. Die beste Antwort hierauf scheint mir das reichhaltige Buch „Psychologische Untersuchung der Konstitution des Kindes“ von *H. Hetzer* zu geben, aber

auch andere Schriften des ehemaligen Wiener Kreises (besonders von *Ch. Bühler*) bringen uns viele solcher Auslegungen. Vorzüglich die erstgenannte Arbeit scheint mir große Errungenschaften hinsichtlich der prägnanten und Wesenszüge erfassenden Testauslegung zu geben. Die Interpretation ist psychologisch angemessen, ja besonders wertvoll und treffend; nur wird zu wenig mit den alten quantitativen Auswertungen aufgeräumt, man läßt die alte Ansicht neben der neuen bestehen, so daß sich die oben zitierten Unstimmigkeiten ergeben.

Freilich soll, dem Plan dieser kritischen Studie entsprechend, auch noch gegenüber den gutgeheißenen Stücken dieser Methode der Persönlichkeitserfassung eine prüfende Besinnung zum Ausdruck kommen. Wir sagen uns: Jede psychologische Persönlichkeitserfassung gründet auf einer vorweggenommenen *Ansicht* vom Menschen und auf der *Idee* von Normen und von Wesentlichem und Nebensächlichem an der Persönlichkeit, die nicht ausschließlich psychologisch, sondern „philosophisch“ sind. Das ist auch bei der Erfassung des Kindes der Fall, möge hier auch das Wort „Persönlichkeit“ nicht die intensive Bedeutung haben wie beim Erwachsenen und mögen die psychologischen und biologischen Bestimmstücke hier einen größeren Geltungsbereich haben als später. Sie beruht doch auch auf philosophischer Anthropologie, da sie wesentliche Charaktereigenschaften und strukturierende Bestimmstücke der Persönlichkeit und des individuellen Lebens an einer Idee vom Menschen zu werten unternimmt; sie kann z. B., um nur einige sozialweltlich naheliegende Gestaltungsfaktoren zu nennen, mehr oder weniger ethisch, leistungsmäßig, philanthropisch, medizinisch usw. eingestellt sein. Sie hat ihre eigene Nomenklatur, sie sieht und unterstreicht, ja gestaltet das eine und vernachläßigt das andere. Dies bildet ein bekanntes Problem, das hier nur erwähnt, aber nicht besprochen werden soll. Wenn wir aber nun die Personerfassung der vorliegenden Studien einen Augenblick aus dieser Perspektive zu überblicken trachten, so kann in Kürze gesagt werden: Die Interpretationen von *H. Hetzer* gestalten eine Erfassung, die das Kind hauptsächlich vom Lebensraum der Schule her sieht. Die Schulanforderungen an Leistungen und die Einordnung in die Schulgemeinschaft (bzw. Kindergartengemeinschaft) modellieren die Kriterien der Charakteristik. (Die Entstehungsgeschichte der Tests scheint dem zu widersprechen, wenn man an die vielen kinderpsychologischen Unterlagen denkt, aber ich kann mich doch nicht dem genannten Eindruck verschließen.) Nicht im Sinne von „ausschließlich“ oder „verfälschend“, auch nicht im Sinne einer Angleichung an Anforderungen einer einseitigen Lernschule, sondern mehr mit der Vorstellung einer Lehr- und Erziehungsanstalt, die schon in vielem der späteren „Schule des Lebens“ gleicht. Mit diesen Vorbehalten dürften meines Erachtens die Wiener Tests in ihrer jetzigen Form und mit der von Frau *Hetzer* gebahnten Auslegung vor allem „Schulreifetests“ sein. Bei den jüngsten Jahr-

gängen kann hiervon natürlich keine Rede sein; gleichwohl scheint dieser „Begriff vom Kinde“ schon sehr früh im Vorschulalter die Auslegung zu beeinflussen. Die Persönlichkeitserfassung in bezug auf Individuelles, spezielle Begabung, bestimmte normale oder abwegige Typen, den ‚endothymen Grund‘, auf das Verhältnis zur persönlichen Welt und Mitwelt, auf die Ausdruckswelt und anderes wird mehr außer acht gelassen. Dagegen wird das Urteil auf den einbezogenen Gebieten ziemlich straff und ausschließend gehalten, so daß doch noch eine gewisse Enge und Schematisierung entsteht. Die scharfe Kritik, die unlängst *H. Asperger* und *J. Feldner*⁷ in ihren „Bemerkungen zu dem Buche ‚Praktische Kinderpsychologie‘ von Prof. *Ch. Bühler*“ (in dem die Wiener Tests eine große Rolle spielen) ausgesprochen haben, kann auch die anhängige Frage beleuchten und — in sehr beschränktem Ausmaße — auch für *H. Hetzers* Buch gelten. In den „Bemerkungen“ heißt es: „Während sich in der Psychologie immer mehr die Erkenntnis Bahn bricht, daß Tests und Statistik für die Erfassung wirklich wesentlicher Züge letztlich unfruchtbar sind, treten uns diese Methoden hier in höchster Übersteigerung entgegen.... Beziehungen zur Welt und zu den Menschen werden unter die Testmaschinerie gebracht und zahlen- und kurvenmäßig dargestellt. Was sich aber beim besten Willen nicht in jene Maschine einspannen läßt, ist für die Verfasserin einfach nicht vorhanden. Schon daraus ergibt sich eine ungeheure Verarmung an psychologischer Erkenntnis.... Immer wieder trifft man in dem Buch auf Pauschalurteile. ... wir sehen (weiter) die Gefahr gegeben, daß unter der Eingeengtheit einseitig angewandter Testmethodik bei Ausschluß medizinischer Gesichtspunkte pathologische Fälle unerkannt bleiben. ... Der in diesem Buch unternommene Versuch, sämtliche Erziehungsschwierigkeiten rein psychologisch aufzulösen, muß daher als ein Rückschritt bezeichnet werden....“

Wenn wir diese Kritik an den genannten Interpretationen *Hetzers* anlegen (für die sie nicht geschrieben ist), dann zeigt es sich sofort, daß sie hier nur sehr eingeschränkt zutrifft. Aber doch scheint mir das „psychologische Weltbild der Testbeurteilung“ auch hier noch teilweise erweiterungsbedürftig und ausbaufähig. Erstens im Sinne einer klaren Bewußtwerdung dessen, daß auch noch bei der Interpretation im Sinne der neueren Arbeiten nur ein begrenztes Persönlichkeitsbild entstehen wird, das uns nicht als *das* Persönlichkeitsbild des Prüflings vorschweben soll. Daß bei der Testuntersuchung allein vieles Wichtige nicht erfaßt wird. Daß sehr viel von den psychologischen, pädagogischen und psychiatrischen Kenntnissen, Gaben und Erfahrungen des Untersuchers abhängt. Und daß vor allem ein umsichtiger Ausbau der nötigen psychopathologischen Interpretation und Anwendung gediegener psychiatrisch-psychologischer Erkenntnisse zu wünschen übrigbleibt, wenn das Verfahren mit seinen Ansprüchen übereinstimmen will.

Es bleibt bei den Untersuchungen mit den Wiener Tests geboten, sein Augenmerk nicht allein auf die testmäßig untersuchten Leistungen und Persönlichkeitseigenschaften zu richten, sondern auch das Testinstrument, seine Maßgeblichkeit und Zulänglichkeit fortwährend kritisch zu beobachten. Man muß immer beide Sachverhalte, Prüflingspsyche und Prüfungsverfahren, anvisieren und einschätzen. Praktisch ergibt sich so: 1. Die Entscheidung soll mehr als bisher bei der aktuellen psychologisch-psychiatrischen Interpretation liegen; diese Interpretation soll noch umsichtiger werden. 2. Wenn ein Prüfling z. B. Tests aus psychologisch nicht sofort durchschaubaren Gründen verfehlt, dann kann uns vielleicht sowohl eine Analyse seiner Gesamtpersönlichkeit als auch eine Auslegung der ganzen Testmaßgeblichkeit und des einzelnen Tests zur richtigen Lösung bringen. 3. Falls ein Verhalten, vor allem ein Versagen, irgendwie psychologisch (psychopathologisch) unwahrscheinlich oder unbegründbar bleibt, so darf es einfach nicht angerechnet werden. Wir versuchen in solchen Fällen, einen oder verschiedene verwandte, gebietsgleiche Tests darzubieten, um der Sache auf den Grund zu kommen. 4. Die einzelnen Tests des Systems sollen noch weiter durchstudiert und verändert werden. Es sollen auch verwandte „Ersatztests“ aufgestellt werden.

Zuletzt: Gerade mit Hilfe einer gewissen Relativierung und Einschränkung der Zuständigkeit des besprochenen Begutachtungsverfahrens, das, wie gesagt, so viel für sich hat, hoffen wir den Ertrag zu fördern.

Schlußbemerkungen.

Zusammenfassend kann ich die Ergebnisse unseres praktischen und theoretischen Studiums der Wiener Tests folgendermaßen umschreiben: 1. Unsere Erfahrungen lassen vermuten, daß man mit den Tests in Amsterdam im großen und ganzen zu ähnlichen Ergebnissen kommt wie bei den bereits untersuchten Kindern anderer Großstädte. 2. Das Testsystem hat große Vorzüge, die seine Erlernung und Anwendung wünschenswert machen. 3. Die Verfasser überschätzen aber die Maßgeblichkeit der Tests, insbesondere die quantitative Bewertung. 4. Der sog. Entwicklungsquotient ist in vielen Fällen kein guter Anzeiger des psychischen Entwicklungsstandes des Kindes. 5. Die Fragen bezüglich der einzelnen Funktionsgebiete oder „Dimensionen“ können nicht so eindeutig und allgemeingültig beantwortet werden, wie es die Bestimmungen des Testverfahrens wahrhaben wollen. Die zahlenmäßige Darstellung von Ergebnissen auf den verschiedenen Gebieten in der jetzigen Form führt zu Fehlurteilen. 6. Das in den letzten Jahren ausgebauten Interpretationsverfahren für die Wiener Tests sollte zu überwiegender Geltung gelangen. Hiermit kann man zur Erfassung wesentlicher Persönlichkeitszüge kommen. 7. Für den weiteren Ausbau dieses

Auslegungsverfahrens (das in seiner heute vorliegenden Form und Nomenklatur den gemachten Ansprüchen nicht voll genügt) empfiehlt sich Bedachtnahme auf die Grenzen testmäßiger Persönlichkeitserfassung, Erweiterung des psychologischen Gesichtskreises, bessere Einbeziehung psychopathologischer Fakten. 8. Der Ausbau kann nicht durch einzelne Anweisungen und Regeln erreicht werden, sondern könnte dadurch entstehen, daß Berufene das Verfahren erproben und es noch weiter „erziehen.“ 9. Es wird versucht, neben Bemerkungen zu einzelnen Tests und kleinen Änderungsvorschlägen einige Richtlinien für eine zu rechtfertigende psychologische Begutachtung mit den Wiener Tests zu geben.

In dieser Studie wird somit hauptsächlich versucht, die rein psychologischen und charakterologischen Qualitäten des Systems zu verstärken und die Testergebnisse so zu plazieren, daß sie in der Bedeutung gesehen werden, die ihnen innerhalb einer vielseitigen Einschätzung des kindlichen Individiums zukommt. Im Sinne dieser Anschauung besteht der Wunsch, daß die Tests nur von gut vorgebildeten und vorbereiteten Versuchsleitern angewendet und verwertet werden. Die in letzter Zeit zunehmende Gewohnheit, einzelne Testgruppen aus dem System ohne die nötigen Erläuterungen abzudrucken (*Weygandt, Dubitscher, Rombouts*), dürfte schädliche Folgen haben. Denn die Wiener Tests verlangen wie alle psychologische Begutachtungsverfahren ein gründliches Studium.

Schriftennachweis.

- ¹ *Bühler, Ch. u. H. Hetzer:* Kleinkindertests. Entwicklungstests vom 1. bis 6. Lebensjahr. Leipzig 1932. — ² *Danzinger, L.:* Der Schulreifetest. Wiener Arbeiten zur Pädagogischen Psychologie, H. 9. Wien 1933. — ³ *Hetzer, H.:* Die seelischen Veränderungen des Kindes bei dem ersten Gestaltwandel. Leipzig 1936. — ⁴ *Hetzer, H.:* Psychologische Untersuchung der Konstitution des Kindes. Leipzig 1938. — ⁵ *Bühler, Ch.:* Praktische Kinderpsychologie. Wien 1938. — ⁶ *Hofstätter, P. R.:* Z. Kinderforsch. 47, 72 (1938). — ⁷ *Asperger, H. u. J. Feldner:* Z. Kinderforsch. 47, 97 (1938). — ⁸ *Wolf, Maria:* Z. Kinderforsch. 44, 191 (1935).
-