

Diagnostische und didaktisch-methodische Funktion von Tests im Unterricht

WERNER LOHMANN

*Klaus-E. Behne (Hg.): Gefühl als Erlebnis - Ausdruck als Sinn. - Laaber: Laaber 1982.
(Musikpädagogische Forschung. Band 3)*

Im Schulfach Musik besteht nach wie vor ein Mangel an zuverlässigen Verfahren der Leistungsmessung. Kaum ein anderes Fach hat der Entwicklung objektiverer Formen für eine Bewertung von Schülerverhalten im Unterricht bisher so wenig Aufmerksamkeit entgegengebracht. Leistungsanforderung, Prüfung und Zensurierung im Fach Musik sind – so Klaus Füller – in der öffentlichen Meinung bis heute verpönt.¹

Andererseits ist Leistungsüberprüfung unverzichtbarer Bestandteil des gegenwärtigen (und wohl auch zukünftigen) Schulwesens. Der Musiklehrer unterliegt der gesetzesmäßigen Verpflichtung, jedem Schüler zweimal jährlich per Zeugnisnote die Qualität der von ihm erbrachten Leistung zu attestieren. Ob er dies widerstrebend ausführt in der Überzeugung, Bewertung durch Festlegen einer Note passe nicht in sein Unterrichtskonzept, oder ob er gerne von der Möglichkeit Gebrauch macht, die Androhung (und Vergabe) einer schlechten Zensur als „disziplinarische Maßnahme“ einzusetzen – in jedem Fall nimmt er eine Festschreibung vor, für die ihm bisher von der Musikdidaktik wenig Hilfestellung geboten worden ist.

Die zu diesem Thema bisher publizierten Beiträge konstatieren entweder die für das Fach Musik besonders unbefriedigende Situation, ohne daraus Konsequenzen zu ziehen, oder stellen Beispiele individueller Leistungsmessung – meist global „Tests“ genannt – vor, ohne daß der Kontext deutlich gemacht wird, in den diese Ansätze integriert werden müßten.

Leistungsmessung bedarf jedoch eines übergeordneten Zusammenhangs, eines integrativen Ansatzes, weil sie verselbständigt eher neue Probleme schaffen als die vorhandenen lösen kann.

Nach meiner Auffassung eignet sich besonders das Konzept der Evaluation dazu, „*Leistungsmessung und Leistungsbewertung in ihrer ganzen Funktionsbreite als integrativen Bestandteil von Lern- und Unterrichtsprozessen theoretisch zu beschreiben und zu begründen*“.²

Der Terminus der Evaluation – in den USA seit langem populär – ist von der Allgemeinen Pädagogik in der Bundesrepublik inzwischen aufgenommen, von der Musikdidaktik jedoch kaum beachtet worden. Allerdings wird er hierzulande häufig in einem zu engen Sinne lediglich als Synonym für Leistungsmessung verwendet. Die Hauptvertreter der Evaluationsforschung (vor

allem Tyler, Cronbach, Seriven sowie Stake) benutzen den Begriff erheblich umfassender, uni „*die Gesamtheit der Vergleichs-, Kontroll- und Bewertungsoperationen zu bezeichnen, die bei der Beurteilung didaktischer Systeme auftreten*“.³ Der damit geschaffene feedback-Prozeß ermöglicht konkrete Ansätze für eine empirische Unterrichtsforschung. Darüber hinaus ergibt sich die Möglichkeit, die vielzitierte Kluft zwischen „Theorie“ und „Praxis“ zu überbrücken, wenn im Rahmen eines Evaluationskonzeptes alle Betroffenen einen ihrem jeweiligen Arbeitsbereich entsprechenden Beitrag leisten. Fachdidaktiker müßten Modelle entwickeln, deren praktische Erprobung Aufgabe der Musiklehrer wäre.

Mir sind im Bereich des Schulfaches Musik bisher nur zwei solche modellartigen Ansätze bekannt.

Zum einen handelt es sich um eine Publikation von Richard Colwell aus dem Jahre 1970, die nur in englischer Sprache vorliegt und darüber hinaus seit längerem nicht mehr käuflich erhältlich ist.⁴ Kurt-Erich Eicke hat seinerzeit einen kurzen Abriß des Inhalts vorgelegt.⁵ Colwells Intention besteht darin, aufzuzeigen, „*how evaluation, properly used, can improve any kind of teaching-learning situation*“.⁶ ‚Richtig benutzt‘ wird dieses Konzept nach Ansicht Colwells dann, wenn es einen regelkreisartigen Charakter hat, so daß das erste „*planning*“ nach Realisierung, Datensammlung und -interpretation von einem erneuten Forschungskreislauf abgelöst wird.

Colwells Arbeit stellt eine Pionierleistung für das Fach Musik dar. Sie ist jedoch auf bundesdeutsche Verhältnisse nur beschränkt übertragbar; zudem verwendet der Autor zur Veranschaulichung seines Konzeptes Beispiele, die außerhalb des Unterrichts vonstatten gehen („*concerts*“, „*contests*“ und „*recording devices*“). Damit wird das Erfassen schulischer Leistungen vor allem im affektiven und psychomotorischen Bereich nicht in der Weise veranschaulicht, wie es sich der Leser des Buches vielleicht erhofft hat.

Die andere Arbeit – Klaus Füllers Dissertation – legt ebenfalls ein Evaluationskonzept zugrunde, das analog dem Curriculum-Prinzip regelkreisartig arbeitet:

- „*Auswahl von Lehrinhalten bezüglich eines definierten Adressatenkreises*
- *Operationale Definition der Lehr- und Lernziele*
- *Herbeiführung von Vermittlungs- und Erhebungssituationen*
- *Erhebung von Daten*
- *Interpretation der Daten*
- *Rückwirkungen auf den Gesamtprozeß*“⁷

Füller verwendet im kognitiven und affektiven Bereich die Arbeiten der Gruppe um Bloom und Krathwohl; für den psychomotorischen Bereich

greift er auf die Taxonomie von E. Simpson zurück. Daraus erarbeitet er fünf Lernzielbereiche für das Schulfach Musik.

Nach eigenen Angaben konnte Füller diesen theoretischen Ansatz selbst nicht umfassend untersuchen; deshalb beschränkte er sich darauf, „*Erfahrungen mit informellen Musikleistungstests zu sammeln*“.⁸

Der Kontext zwischen Leistungsmessung in Form von Tests und dem Evaluationskonzept wird hier deutlich: propagiert wird der Einsatz eines Tests als objektivierte Verfahren innerhalb der empirischen Unterrichtsforschung. An einen derartigen Test ist eine Reihe von Anforderungen zu stellen, denen die o. a. Einzelansätze mit gleichem Namen nicht genügen können:

- Einsatzmöglichkeit für alle Schüler, auch wenn sie nicht über (vor allem außerschulische) Vorerfahrung verfügen;
- Hervorgehen aus dem Unterricht; d. h. Überprüfung von Lernprozessen;
- objektivierte, problemlose Durchführbarkeit ohne eine spezielle Ausbildung als Voraussetzung;
- Entscheidungshilfe für weitere Überlegungen bezüglich Inhalten und Methoden des Musikunterrichts;
- Einbindung in einen curricularen Zusammenhang als Bestandteil der Evaluation.

Heute kann der Musiklehrer theoretisch fünf unterschiedliche Arten von „Tests“ im Unterricht einsetzen (womit deutlich wird, wie vage die pauschale Angabe „Musiktest“ im Grunde ist):

1. einen standardisierten Musikalitätstest;
2. einen standardisierten Musikleistungstest;
3. einen unterschiedlich konzipierten „Zetteltest“;
4. einen informellen kriteriumsorientierten Test;
5. einen informellen lernzielorientierten Test.⁹

zu 1.

Wenngleich die Thematik „Musikalität“ und „Musikalitätsmessung“ bei uns häufig (bisweilen auch heftig) diskutiert wird, beschränkt sich der Einsatz von Meßverfahren in der Praxis auf zwei Tests englischsprachigen Ursprungs, deren deutsche Ausgaben 1968 (Arnold Bentley) bzw. 1966 (Carl E. Seashore) erschienen sind. Die vielfach erhobenen Bedenken gegen Tests zur Messung musikalischer Begabung gelten für diese beiden Verfahren ebenso wie für andere, weniger verbreitete:

Zunächst herrscht noch immer Uneinheitlichkeit sowohl in Bezug auf den allgemeinen als auch auf den musikalischen Begabungsbegriff hinsichtlich ihrer Bedeutung. Einerseits hat dies sprachliche (vor allem Übersetzungs-

technische) Gründe. Andererseits formulieren manche Testautoren eher eine persönliche Überzeugung als eine wissenschaftliche Argumentation.

Ungeklärt sind ebenfalls die beiden großen Kontroversen,

- a) ob primär Anlage oder primär Umwelt (= Lernprozesse) für das Zustandekommen der Musikalität verantwortlich sind, und
- b) ob sich das Wesen musikalischer Begabung über isolierte Elemente oder über geschlossene, ganzheitliche Formen erfassen läßt.

Dabei bedeutet eine Festlegung in der einen oder anderen Richtung zugleich auch eine (bildungs)politische Entscheidung.

Aus testpsychologischer Sicht ist zu monieren, daß die Standardisierungen im Vergleich zu Tests in anderen Bereichen den Hauptgütekriterien nur eingeschränkt entsprechen.

Aus der Sicht der Allgemeinbildenden Schule ergeben sich darüber hinaus Bedenken gegen den Grundgedanken des Musikalitäts-Konzeptes, weil es Anliegen der Schule sein muß, allen Schülern eine Förderung im Musikunterricht zukommen zu lassen. Die normativen Begriffe und das Wertdenken dieses Konzeptes sind nicht mit der Idee von Chancengleichheit in Einklang zu bringen.

*„Würde man die Betrachtung von Musikalitätstests noch weiter ausführen“ – so Helga de la Motte-Haber –, „so würde man immer wieder denselben Sachverhalt hervorkehren müssen, nämlich, daß es kein geeignetes Prüfverfahren für eine quantitative Bestimmung der musikalischen Begabung gibt.“*¹⁰

zu 2.

Standardisierte Musikleistungstests wird man in einschlägigen Übersichten selten finden. Der mit einer Standardisierung verbundene Aufwand steht offenbar in einem Mißverhältnis zu der Möglichkeit, eine bestimmte Fähigkeit bezüglich eines begrenzten Themas in Relation zu den Leistungen einer repräsentativen Stichprobe aus der Gesamtpopulation zu testen.

Eine gewisse Bekanntheit genießt lediglich Robert Wagners Projekt aus den Jahren 1963-1968, dessen genauer Titel aufmerken läßt, weil die Polarisierung Musikalitätstest – Leistungstest aufgehoben scheint: *„Untersuchungen zur Entwicklung der Musikalität mit Hilfe eines Musikleistungstests“*. Eine detaillierte Betrachtung vermittelt den Eindruck, daß – wie bei anderen Tests im Bereich Musik – die Einhaltung der wichtigsten Gütekriterien nicht gewährleistet erscheint. Auch bleibt Wagners Verständnis des Musikalitätsbegriffs für den Leser diffus.

Wichtiger erscheinen jedoch die grundsätzlichen Einwände, die in den vergan-

genen Jahren gegen den Einsatz standardisierter Testverfahren erhoben worden sind. Auf die Kritik an der mathematischen Axiomatik der klassischen Testtheorie kann hier nicht eingegangen werden. Inhaltlich wird vor allem geltend gemacht:

- standardisierte Tests sind nicht flexibel genug, um angemessenen Lernfortschritte erfassen zu können;
- häufig fehlt die Übereinstimmung zwischen Lehr- und Lernzielen des Lehrers und des Tests;
- einziges Ziel ist die Differenzierung zwischen den Meßobjekten;
- das Verfahren ist zu sehr technisch (statistisch an der Fehlertheorie) orientiert; es fehlt eine differenzierte psychologische Fundierung.¹¹

Daß standardisierte Verfahren selten im Musikunterricht Verwendung finden, liegt wahrscheinlich ebenso an der geringen Zahl käuflich zu erwerbender Tests als auch an einer gewissen Unsicherheit gegenüber dem statistischen Bereich (vor allem bei der Daten-Interpretation), die gerade bei Musiklehrern eine Rolle spielen dürfte. Viel häufiger verwenden diese eine Art von Tests, die in keinem Handbuch zu finden ist:

zu 3.

Meist wird von sog. „Zetteltests“ gesprochen; es gibt aber eine Vielzahl weiterer Begriffe vom „Kurztest“ bis zur „schriftlichen Überprüfung der Hausaufgabe“. Nach der Rechtslage ist eine derartige Leistungsüberprüfung in NRW beispielsweise durchaus erlaubt. Durften die „Zetteltests“ früher höchstens mit einer Punktzahl versehen werden, kann der Lehrer heute auch eine Zensur darunter setzen. Das Gesetz gesteht diesen Tests jedoch nur eine untergeordnete Funktion zu:

„Neben den vorgeschriebenen schriftlichen Arbeiten zur Leistungsfeststellung (die in der Sekundarstufe I im Fach Musik entfallen; d. Verf.) sind in allen Fächern gelegentliche kurze schriftliche Übungen zulässig. Sie dürfen sich nur auf begrenzte Stoffbereiche in unmittelbarem Zusammenhang mit dem jeweiligen Unterricht beziehen und können wie eine zusätzliche mündliche Leistung bewertet werden; die Überprüfung der mündlichen Leistung darf dadurch nicht ersetzt werden.“¹²

Die Praxis sieht anders aus: die „kurzen“ Tests erstrecken sich häufig über zwei Schulstunden; anstelle eines „unmittelbaren Zusammenhangs“ wird der Stoff eines Viertel- oder gar 'Halbjahres geprüft; aus der „zusätzlichen Leistung“ wird die (allein) entscheidende Grundlage der Zeugniszensur. Es ist anzunehmen, daß die meisten Lehrer angesichts des Mangels an anderen Möglichkeiten einer exakten Leistungserfassung aus der Not eine Tugend zu

machen versuchen. Jedoch wird damit weder den Schülern noch dem Fach Musik ein Gefallen erwiesen. Um so dringender erhebt sich die Forderung, dem Lehrer geeignetere Verfahren an die Hand zu geben, die ihm eine objektivierte Messung erlauben. Deutlich wird, daß weder aufwendige und komplizierte standardisierte Musiktests noch in einer Freistunde „ausgedachte“ Lehrerfragen, die den Begriff „Test“ in Anspruch nehmen, geeignet sind, die Frustration des Musiklehrers bezüglich seiner Leistungsmessung und Zensurengebung zu mindern. (Es stimmt bedenklich, wenn sogar erfahrene Didaktiker inzwischen von der „*Not der Musiklehrer*“, von der „*leider unumgänglichen Zeugnisnote*“ glauben sprechen zu müssen.¹⁴)

So bleiben die Verfahren übrig, die unter den Ziffern 4. und 5. aufgelistet wurden und denen beiden der Anspruch gemeinsam ist, objektiv (bzw. objektiviert) und informell zu sein. Sie unterscheiden sich in einem teststatistischen Punkt, der allerdings didaktische Konsequenzen hat.

zu 4.

Der Begriff „*kriteriumbezogen*“ wurde von Glaser in die deutsche Literatur eingebracht, um diese Testart von den „*normbezogenen*“ Tests zu unterscheiden. Dem Verfasser eines kriteriumorientierten Tests geht es ausschließlich darum, festzustellen, ob die untersuchte Gruppe das von ihm erstellte Kriterium erreicht hat. Hierzu werden Bewertungsschlüssel wie z. B. das „*90/90-Kriterium*“ erstellt: der Test ist brauchbar und das Ziel erreicht, wenn mindestens 90 Prozent der Schüler 90 Prozent der Aufgaben richtig gelöst haben.

Die Vertreter dieses Ansatzes lehnen eine statistische Aufbereitung der Tests ab. Da sie die Vorbedingung stellen, daß das „*Kriterium*“ von nahezu allen Schülern erfüllt wird, ist eine Normalverteilung ausgeschlossen, die wiederum die Voraussetzung für die Anwendung der klassischen Testtheorie darstellt. Sieht man von diesem statistischen Problem, mit dem sich die Experten seit längerem intensiv auseinandersetzen, ab, bleiben inhaltliche Bedenken vor allem aus praktischer Sicht. Es erscheint schwer vorstellbar, einen Test zu konstruieren, der einerseits alle wesentlichen Ereignisse des Unterrichts erfaßt und die dabei vermittelten Fähigkeiten überprüft und andererseits von fast allen Schülern fehlerfrei bearbeitet werden kann. Auch wiegt die fehlende Möglichkeit schwer, mit den bisher zur Verfügung gestellten Verfahren die Qualität eines derartigen vom Lehrer erstellten Tests zu überprüfen. Offen bleibt nämlich stets, welcher Faktor letztlich für das Resultat verantwortlich gewesen ist – unabhängig davon, ob das Kriterium erreicht oder nicht erreicht wurde.

Darüber hinaus kann ein kriteriumorientierter Test in unserem jetzigen Schulsystem nicht dazu beitragen, die Zensurfindung im Musikunterricht zu erleichtern, weil ein „guter“ Test eine Klasse voraussetzt, die nur aus „guten“ Schülern besteht. Daraus jedoch eine Einheitszensur abzuleiten, erbringt nicht nur Schwierigkeiten mit der Schulaufsichtsbehörde, sondern ist auch in der Sache ungerechtfertigt: die Lernprozesse, die Musikunterricht hervorrufen soll, werden in der Regel nicht bei allen Schülern in gleicher Weise erfolgreich verlaufen.

zu 5.

Der im folgenden skizzierte Typ eines lern- bzw. lehrzielorientierten Tests (die Unterscheidung ist in diesem Zusammenhang weniger von Bedeutung) wird bewußt gegen die vorausgegangene Form abgesetzt. Manche Autoren verwenden die Arten 4. und 5. synonym. Mir scheint aber eine Differenzierung sinnvoll, weil der Informelle lernzielorientierte Test a) keine fixierte Qualität des Resultats voraussetzt und b) aufgrund der Streuung der Testwerte eine Datenaufbereitung zuläßt.

Innerhalb eines Evaluationskonzeptes vermag dieser Testtyp die Leistung einer Schülergruppe in einer allen anderen Verfahren überlegenen Weise zu erfassen. Er stellt Relationen zwischen Zielen und Methoden her, ohne formell durch Anwendung abstrakter pädagogischer Gütemaßnahmen zu Entscheidungen zu führen.¹⁴ Lernzielorientierte Tests *„wollen keine Aussage über ‚die‘ Leistung von Schülern einer bestimmten Altersstufe in einem bestimmten thematischen, Bereich für alle Zeiten machen, sondern lediglich ermitteln, welche Leistungen an den – zugegeben: von den jeweiligen Lehrern ausgewählten – Inhalten von Schülern des engen Wirkungsbereichs einer oder mehrerer Schulen erbracht werden, bei intersubjektiv konsistentem Beurteilungsmaßstab“.*¹⁵

Der lernzielorientierte Test impliziert bereits durch seinen Namen einen Kontext zwischen Planung von Unterricht und Erfassen der durch Unterricht möglich gewordenen Leistung. Er verhindert damit eine isolierte Messung und zwingt den Lehrer zu reflektiertem Handeln. Im Unterschied zu anderen Testverfahren hat er so eine zweifache Funktion: zum einen eine diagnostische, zum anderen eine didaktisch-methodische.

Die diagnostische Funktion wird in der Regel als „die“ Leistung eines Tests apostrophiert. Sie ist in der einschlägigen Literatur und in den Handreichungen zu standardisierten Tests Gegenstand ausführlicher Erörterungen. Der gesamte statistische Bereich dient dazu, die Exaktheit der Diagnose zu gewährleisten. Auch lernzielorientierte Tests weisen diesen diagnostischen As-

pekt auf einschließlich einer – allerdings im Vergleich zu standardisierten Verfahren abgewandelten (s. u.) – Prüfstatistik; sollen sie doch über das Resultat des Unterrichtsgeschehens verlässlich Auskunft geben.

Die Diagnose bleibt hier jedoch nicht auf das Messen individueller Schülerleistungen beschränkt: geht man (etwa im Sinne der Themenzentrierten Interaktion Ruth Cohns) davon aus, daß das Ergebnis von Unterricht aus den Wechselbeziehungen der Triade Lehrer – Schüler – Thema resultiert, wird deutlich, daß eine univariate Betrachtung des bzw. der Schüler(s) im Bereich Leistung diese Interaktionen, die permanent erfolgen, zu Unrecht vernachlässigt.

Schon vor mehr als zehn Jahren forderten Gaude und Teschner: „*Die Rohdaten aus den Informellen Tests bedürfen einer zweifachen Weiterverarbeitung, der statistischen und der didaktischen.*“¹⁶ Jedoch wird der pädagogische Kontext in Veröffentlichungen, die sich mit Tests in der Schulpraxis beschäftigen, selten berücksichtigt und eine didaktische Fragestellung zugunsten statistischer Gütekriterien vernachlässigt.

Bei standardisierten Schulleistungstests muß beispielsweise die Einbindung in curriculare Konzepte fehlen bzw. gering sein, weil einmal die Testnormen auf einer Jahre zuvor vorgenommenen Eichung beruhen und zum anderen die zwingende Möglichkeit, einen solchen Test jederzeit an jedem Ort einsetzen zu können, eine Anlehnung an ein konkretes •Curriculum nicht zuläßt.

Gronlund hat in einer grundlegenden Arbeit zum Thema „Leistungstests“ die ihm am wichtigsten erscheinenden „*Grundprinzipien der Leistungsmessung*“ zusammengefaßt.¹⁷ Die sechs Thesen postulieren eine testtheoretisch einwandfreie Erhebung und Auswertung der Daten. Sie formulieren jedoch gleichzeitig den didaktischen Zusammenhang, für den diese statistische Exaktheit letztlich nur relevant ist:

1. Leistungstests sollten klar definierte, mit den Lernzielen in Einklang stehende Lernergebnisse messen.
 2. Leistungstests sollten eine repräsentative Stichprobe der Lernergebnisse und des im Unterricht behandelten Stoffes messen.
 3. Leistungstests sollten die zur Messung der Lernergebnisse jeweils geeigneten Testaufgabentypen enthalten.
 4. Leistungstests sollten auf den jeweiligen Verwendungszweck der Ergebnisse zugeschnitten werden.
 5. Leistungstests sollten so zuverlässig wie möglich sein und mit Vorsicht gedeutet werden.
 6. Leistungstests sollten dazu dienen, den Lernvorgang zu verbessern.
- Gelingt es, die in diesen Forderungen enthaltene Verbindung diagnostischer

und didaktischer Aspekte in die Praxis des Musikunterrichts umzusetzen, dann könnte ein derartiges Testverfahren die Probleme der Leistungsmessung hier wesentlich verringern. Deshalb ist es notwendig, daß gerade im Fach Musik ein didaktisches Konzept von vornherein die Integration von Informellen lernzielorientierten Tests vorsieht. Sigrid Abel-Struth hat darauf hingewiesen, daß vor allem durch „*Verfahren der Evaluation*“ die fachspezifischen Aufgaben „*Aufhellung des musikalischen Lernprozesses*“ und „*Bedingungen des musikalischen Lernens*“ angegangen werden können und dadurch „*mehr Erkenntnis über die Effektivität von Lernvorgängen in der Schule*“ ermöglicht wird.¹⁸

Werden bei der Konzeption von Tests über diagnostische Fragen hinaus auch didaktisch-methodische Überlegungen angestellt, dann bedarf das schulische Umfeld, in dem ein Test eingesetzt wird, einer weitergehenden Reflexion als bisher üblich. Einige Beispiele notwendiger Überlegungen:

- Ein Test darf nicht primär Quelle von Schülerangst sein. Die eher kognitive Ausrichtung von Leistungstests könnte dazu führen, affektives Geschehen im Schüler zu vernachlässigen. Entsprechend muß der Lehrer sowohl das Testverfahren selbst als auch die Bedeutung, die er den Ergebnissen beimißt, ausführlich erläutern; er muß das Phänomen der Testangst thematisieren, weil nicht erwartet werden kann, daß gerade ein ängstlicher Schüler dieses Problem von sich aus zur Sprache bringt.
- Ein Test sollte weder Ausgangs- noch Endpunkt von Unterricht sein. Testverfahren in der ersten Stunde, „um einmal zu sehen, wie die Klasse ist“, sind pädagogischer Nonsens; Tests, die kommentarlos (und das heißt auch: von seiten der Schüler her ohne die Möglichkeit einer Stellungnahme) eine Unterrichtsreihe beenden, werden in erster Linie eine diffuse Unsicherheit fördern und können kaum zur Verbesserung des Verhältnisses Lehrer – Klasse beitragen.
- Ein Test sollte vom Lehrer nicht nur daraufhin interpretiert werden, welche Fehler einzelne Schüler gemacht haben, sondern auch, wo sich – aus den Resultaten ersichtlich – Schwachstellen seines eigenen Unterrichts zeigen. Seine Autorität wird nicht sinken, wenn er die Schüler über diesen weiteren Nutzen eines Tests informiert.
- Ein Test sollte von den Schülern als ein Verfahren erkannt (und auch anerkannt) werden, das besser als die herkömmlichen Methoden geeignet ist, ihre tatsächliche Leistung zu erfassen. Dadurch kann den Schülern deutlich gemacht werden, daß die. (richtige) Verwendung eines Tests letztlich auch in ihrem Interesse liegt.
- Ein Test sollte auch als methodischer Bestandteil des Unterrichts gese

hen werden. Das Methodenproblem ist in der didaktischen Diskussion der letzten Jahre ein wenig in den Hintergrund getreten. Dabei gibt es hier zahlreiche Fragestellungen, die noch der klärenden Untersuchung bedürfen. So wäre z. B. der Stellenwert einer fast ausschließlich auditiven Erarbeitung von Musikwerken im Unterricht zu hinterfragen. Es ist nämlich keineswegs so, daß alle analytischen Aufgabenstellungen für die Schüler am ehesten über das Gehör zu lösen sind. Ein Testdesign mit mehreren Experimentalgruppen könnte darüber Auskunft geben, welcher Aspekt der musikalischen Analyse eher auditiv, welcher audiovisuell oder welcher auf rein visuellem Wege optimal von den Schülern erfaßt werden kann.¹⁹

Betrachtet man einen Test auch als Unterrichtsmethode, wird deutlich, daß er nicht dem Anspruch genügen kann, Unterrichtsgeschehen in toto zu erfassen. Der Verzicht auf diese potentielle Einschränkung fällt allerdings nicht schwer, sofern man stattdessen den Test als Unterrichtsimpuls versteht; sofern es gelingt, hierdurch bei den Schülern eine Motivationsquelle zu erschließen.

Fehlen derartige Überlegungen, können dadurch die Vorteile eines lernzielorientierten Tests rasch zunichte gemacht werden. Man sieht, daß die Beschreibung der diagnostischen Funktion, wie sie in allen einschlägigen Lehrbüchern ausgiebig erfolgt, nicht ausreicht, wenn die didaktisch-methodische Komponente fehlt. Deutlich wird auch, daß das Primat mathematisch-statistischer Überlegungen in der Schule keine Gültigkeit haben darf.

Es erscheint angebracht, an einem Beispiel zu veranschaulichen, wie die geschilderte Konzeption eines Informellen lernzielorientierten Tests als Bestandteil der Evaluation von Musikunterricht realisiert werden kann.

Dazu wird im folgenden kurz über eine entsprechende Untersuchung berichtet, die ich im Frühsommer 1981 an einem Oberhausener Gymnasium durchgeführt habe.

Beteiligt waren zwei 6. Klassen und je ein Kurs der Jahrgangsstufen 9 und 12. Zur Gewährleistung einer besseren Übersicht wurde für alle Versuchsgruppen eine gemeinsame Zielsetzung zugrunde gelegt:

„Die Schiller sollen ostinate Formen in der Musik als Synthese von Wiederholung und Veränderung, von Freiheit und Bindung erkennen und erfahren über hörende Wahrnehmung, über verschiedene Notationsformen, über eigene Realisationen und durch Analysen.“

Zentrale Forderung war die Einheit von Lernzielen, Unterrichtsgeschehen und Testinhalten, denn nur auf dieser Basis ist Evaluation möglich.

Der Unterricht wurde von zwei Kolleginnen erteilt, die – nebst einem auch in

der Curriculum-Entwicklung tätigen Fachleiter – ebenfalls an der Entwicklung der Unterrichtsstunden beteiligt waren und mit mir die Lernziele und die Testaufgaben erörterten, um so eine Kontentvalidität zu gewährleisten.

Die Zielsetzung war unter dem Gesichtspunkt ausgewählt worden, daß alle Bereiche von Schülerverhalten im Unterricht vertreten sein sollten, wobei eine altersspezifische Differenzierung vorzunehmen war. Mangels eines einheitlichen Curriculum konnte die Zielsetzung nicht in eine übergeordnete Gesamtkonzeption eingebracht werden, so daß sie eine persönliche Entscheidung darstellt. Interessant erschien mir vor allem die Möglichkeit, sich von der Beschränkung auf eine bestimmte Epoche, eine bestimmte Art von Musik oder (meist daraus resultierend) auf eine oder nur wenige Formen von Unterricht frei zu machen. (Diese in neueren Curricula geforderte Offenheit findet man in den Unterrichtswerken bisher noch selten wieder.) Die Formulierung des Themas kann altersspezifisch modifiziert werden; etwa

- „verschiedene Musikteile – vergleichbare Form" in 6;
- „Ostinato-Formen in alter und neuer Musik" in 9; oder
- „Freiheit und Bindung – Komposition und Improvisation mit einer gegebenen Grundformel" in 12.

Auf eine Unterteilung der Lernziele in die psychischen Qualitäten „kognitiv, affektiv, psychomotorisch" wurde bewußt verzichtet, weil das enge Zusammenwirken dieser Aktivitäten im Unterricht eine Trennung nicht zuläßt.²⁰

Für die Unterrichtsreihen, die sich bis zur Testdurchführung je nach Jahrgangsstufe über sieben bis elf Stunden erstreckten, wurden elf Lernziele formuliert, die für alle Klassen- Gültigkeit haben sollten. Darüber hinaus wurden ergänzend jahgangsspezifische Ziele intendiert.²¹

Der Einstieg erfolgte altersentsprechend in unterschiedlicher Weise; z. B. über eigene Begleitung zu einem neu erlernten Lied in Jggst. 6 oder über eine Orgelbesichtigung in Jggst. 9, bei der den Schülern eine Passacaglia vorgespielt wurde, die anschließend im Unterricht aufgegriffen wurde.

An die Unterrichtsstunde, in der der Test vorgelegt wurde (Testdauer in allen Klassen 45 Minuten), schloß sich eine Stunde zur Besprechung der Testaufgaben und der Testdurchführung an; danach wurde die Unterrichtsreihe mit weiteren Beispielen fortgesetzt.

Die Tests sollten über die schon an anderer Stelle beschriebenen Bedingungen hinaus vier weitere Forderungen erfüllen:

- sie dürfen sich nicht auf das Abfragen isolierter Fakten beschränken, was oft als ihre einzige Fähigkeit dargestellt wird;
- sie müssen die Hauptgütekriterien (s. u.) gewährleisten und deshalb soziometrisch einwandfrei konzipiert und ausgewertet werden;

- sie müssen quasi eine Operationalisierung der Lernziele des vorangegangenen Unterrichts darstellen;
- sie müssen Höraufgaben enthalten, wenn sie Unterricht mehr als punktuell widerspiegeln wollen und Musikhören eine wichtige Rolle in den vorausgegangenen Stunden gespielt hat.

Vor allem der letzte Punkt — bisher in Informellen Tests kaum realisiert — hat nach meiner Ansicht zentrale Bedeutung. Mir ist keine ernsthaft diskutierte didaktische Richtung der Gegenwart bekannt, die auf Musikhören (wie immer dieser Begriff im einzelnen verstanden wird) als wesentlichen Bestandteil des Unterrichts verzichtet. Zwingende Konsequenz ist es dann, bei der Überprüfung eines so gestalteten Unterrichts ebenfalls auf Höraufgaben zurückzugreifen, auch wenn der mit dem Erstellen der Hörbeispiele verbundene Mehraufwand zunächst als Hemmschwelle erscheinen könnte.

In den bei dieser Untersuchung verwendeten Tests wurden sieben der 20 Punkte für Hörleistungen vergeben. Alle Höraufgaben wurden bei Melodiebeispielen vom Klavier, bei Rhythmusbeispielen vom Holzblock auf Tonband aufgezeichnet und im Test abgespielt, wobei auf gleiche Aussteuerung und gleiche akustische Verhältnisse geachtet wurde, um den Einfluß intervenierender Variablen möglichst gering zu halten.

Der Test wurde allen Klassen zuvor angekündigt, weil mir der von manchen Lehrern geschätzte „Überraschungseffekt“ sehr fragwürdig erscheint.

Als Aufgabentypen wurden gewählt:

- a) freie Kurzantwort; b) Mehrfachwahl-Aufgaben (multiple choice) und c) Zuordnungsaufgaben.

Die folgenden fünf Aufgaben, die als typische Beispiele ausgewählt wurden, entstammen den Tests der Klasse 6.

1. Du hörst fünf kurze Klavierstücke aus dem Bereich des Jazz. Von diesen fünf Musikstücken haben zwei ein Ostinato in der Baßstimme. Kreuze also nach dem Hören zwei Ziffern – die Nummern der beiden Stücke mit einem Ostinato – an:

Nr. 1 Nr. 2 Nr. 3 Nr. 4 Nr. 5

2. In dem folgenden Hörbeispiel wird auf dem Klavier in der rechten Hand über einer Bluesformel improvisiert. Worin besteht das Ungewöhnliche an dieser Improvisation:

- sie benutzt nur zwei verschiedene Notenwerte
- sie benutzt nur zwei verschiedene Tonhöhen
- sie benutzt nur drei verschiedene Tonhöhen
- sie benutzt alle Töne der Tonleiter

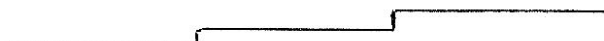
3. Du siehst unter diesem Text drei graphische Darstellungen. Es sollen

gezeichnete Ostinato-Motive sein. Dazu hörst Du gleich vier Musikbeispiele. Es sind die drei abgebildeten Graphiken und noch ein viertes Ostinato, Aas aber nicht graphisch dargestellt ist. Du sollt nun herausfinden, welche Graphik zu welchem Klangbeispiel gehört. Schreibe also nach dem Hören vor jede Zeichnung die Nummer, unter der sie vorgespielt wurde. Bedenke: für ein Hörbeispiel gibt es keine dazugehörige Graphik!

Nr.



Nr.



Nr.



4. Du hast im Unterricht eine Passacaglia und einen Blues kennengelernt, denen beiden ein Ostinato zugrunde lag. Welche der folgenden Aussagen über einen wesentlichen Unterschied zwischen diesen beiden Stücken ist richtig?

- Bei der Passacaglia bleiben die Oberstimmen gleich, beim Blues ändern sie sich.
- Bei der Passacaglia bleibt das Ostinato immer gleich, beim Blues ändert es sich regelmäßig.
- Bei der Passacaglia kann der Spieler einzelne Töne verändern, beim Blues darf er das nicht.
- Bei der Passacaglia sind alle Stimmen notiert, beim Blues werden Stimmen improvisiert.

5. Wann ist ein Thema als Ostinato besonders gut geeignet?

- wenn es keine Vorzeichen hat
- wenn es einen 4/4-Takt hat
- wenn man es leicht wiedererkennen kann
- wenn es sehr lang ist

Die Testsituation wurde weitgehend standardisiert: nicht nur alle Aufgaben einschließlich der Hinweise und Lösungsmöglichkeiten waren auf den ausgegebenen Blättern ausformuliert; auch die Testanweisung wurde von der Einführung in die Testsituation bis zum abschließenden Verweis auf die Besprechung in der folgenden Stunde für alle Schüler gleich schriftlich fixiert vorgegeben (hinzu kamen, wie erwähnt, die vorbereiteten Tonbandbeispiele). Damit wird

die Testdurchführung unabhängig vom jeweiligen Fachlehrer, was für die Vergleichbarkeit der Ergebnisse notwendig ist und die Durchführungsobjektivität gewährleistet.

Entsprechend der Aufgliederung in Lernziele für alle oder nur für eine Gruppe enthalten die Tests Aufgaben, die (jeweils altersentsprechend abgewandelt) allen Kursen gestellt, und solche, die speziell für eine Versuchsgruppe entworfen wurden. Allgemein kann der Unterschied zwischen Tests zum gleichen Thema für verschiedene Jahrgangsstufen bestehen

- in der unterschiedlichen Testlänge;
- in der altersangemessenen Formulierung der Testanweisungen und Aufgabenstellungen;
- in der Wahl zunehmend ähnlicher Falschlösungen bei Mehrfachwahlaufgaben für ältere Schüler, von denen größere Differenzierungsfähigkeit erwartet wird;
- in der Erweiterung der Aufgabenstellung vor allem hinsichtlich produktiver Leistungen zu dem dargebotenen musikalischen Material in höheren Klassen.

Vor einer Darstellung der wichtigsten Ergebnisse ist ein kurzer testtheoretischer Exkurs notwendig.

Zur Zeit herrscht Uneinigkeit bezüglich der Frage, ob überhaupt und ggfs. nach welchen Modellen Informelle Tests statistisch aufbereitet werden sollen. Ich halte eine entsprechende Analyse für unabdingbar, wenn die Leistungsmessung mit Hilfe von Tests der bisherigen Praxis nachhaltig überlegen sein soll. Andererseits scheidet die Verwendung der vollständigen klassischen Testtheorie aus.

So plädiere ich für die Beibehaltung aller wichtigen Testgütekriterien, jedoch unter Verwendung vereinfachter Berechnungsverfahren, deren Anwendung auch einem statistisch ungeübten Musiklehrer in kürzerer Zeit vertraut gemacht werden kann.

Die Berechnung folgender Größen erscheint mir angebracht:

- Mittelwert
- Standardabweichung
- Reliabilitätskoeffizient
- Standardmeßfehler
- zusätzlich eine Aufgabenanalyse bezüglich Schwierigkeit, Trennschärfe und Distraktorenqualität.

Die Objektivität kann bezüglich der Durchführung durch eine wie oben beschriebene Anlage der Versuchssituation und bezüglich der Auswertung durch eine zuvor erfolgte eindeutige Festlegung der Bewertung gesichert werden.

Die Validität läßt sich bei lernzielorientierten Tests nicht formelmäßig berechnen, sondern nur als Kontent-Validität durch das Rating-Urteil von Fachkräften erzielen, was eine gewisse subjektive Komponente nicht ausschließen läßt.

Vereinfachte Formeln für Gütekriterien bei Informellen Tests finden sich in der einschlägigen Literatur."

An die Zuverlässigkeit eines Informellen Tests sind geringere Anforderungen zu stellen als an ein entsprechendes standardisiertes Verfahren. So werden Werte oberhalb von .50 als brauchbar, solche über .60 als voll zufriedenstellend bezeichnet. Nach meinen bisherigen Erfahrungen ist es unrealistisch, auf wesentlich höhere Koeffizienten zu hoffen.

Der 1. Test in einer 6. Klasse erbrachte bei einer Standardabweichung von 1,97 und einem Mittelwert von 16,76 (von 20) Punkten eine Reliabilität von nur .32; also einen unakzeptabel niedrigen Wert. So wurde eine Testrevison erforderlich, bei der die aufgrund der Aufgabenanalyse gefundenen Schwachstellen zu eliminieren waren.

Eine Erhöhung der Reliabilität läßt sich über Verbesserung von Schwierigkeit und Trennschärfe und/oder Verlängerung des Tests erzielen. Letztere Möglichkeit wurde wegen der Vergleichbarkeit der beiden Parallelklassen nicht in Anspruch genommen, zumal selbst eine Verdoppelung der Aufgabenzahl – praktisch kaum durchzuführen – die Reliabilität höchstens um .10 bis .15 ansteigen läßt.

Für die revidierte Fassung wurden 40 Prozent der Aufgaben überarbeitet bzw. ausgetauscht. Sodann wurde die Testrevison einer zweiten 6. Klasse vorgelegt, in der zeitversetzt dieselbe Kollegin die gleiche Unterrichtsreihe durchgeführt hatte. Die Streuung fiel dabei größer aus ($s = 2,82$ bei $M = 13,79$), wodurch der Reliabilitätskoeffizient erheblich stieg: von .32 in der ersten auf .48 in der zweiten Fassung. Bessert man nun noch die restlichen Schwachstellen aus, ist die Verlässlichkeit deutlich über die .50-Grenze zu heben.

Die Tests in den beiden anderen Jahrgangsstufen erbrachten trotz vergleichbarer Zielsetzung, Verlaufsplanung und Testkonzeption erstaunlich abweichende Ergebnisse: wies der Test für die Jggst. 9 praktisch keine Reliabilität auf, erreichte bereits die erste Fassung für die Jggst. 12 den sehr beachtlichen Wert von .58.

Auf Einzelheiten der Testanalyse kann hier nicht eingegangen werden. Die wichtigste Erfahrung war jedoch, daß der Konzeption eines Informellen Tests in der Regel mehrere Revisionen aufgrund der Datenanalyse folgen müssen, ehe eine akzeptable und wiederholt verwendbare Endform vorliegt.

Dabei sind drei Tendenzen der Analyse denkbar, deren erste den häufigsten Fall darstellt:

1. Der Test ist grundsätzlich brauchbar, seine Verlässlichkeit reicht jedoch nicht aus, sondern muß durch gründliche Aufarbeitung auf ein akzeptables Mindestmaß erhöht werden.
2. Der Test weist grundsätzliche Mängel auf, die über Schwachstellen im Detail hinausreichen. Eine zufriedenstellende Verbesserung ist kaum möglich, so daß der Versuch einer Revision nicht sinnvoll erscheint. Der Test ist unbrauchbar.
3. Der Test weist bereits in der ersten Fassung eine so große Stimmigkeit auf, daß die erzielten Werte vollauf genügen. Man kann ihn zunächst unverändert wiederverwenden, um später kleinere Unstimmigkeiten auszubessern.

Da auch bei sorgfältigster Planung nicht vorherzusehen ist, welche der drei Möglichkeiten im Einzelfall zutrifft, wird somit deutlich, daß ein Verzicht auf die Aufbereitung der Daten nicht zulässig sein kann.

Es wurde bereits festgestellt, daß ein Test innerhalb eines umfassenderen Konzeptes von Unterrichtsforschung nicht Endpunkt des Geschehens ist und weiterhin im Sinne einer Mitwirkung aller Beteiligten auch die Meinung der Schüler einbezogen werden muß. Entsprechend folgte in den jeweils nächsten Stunden eine Testerörterung, wobei in erster Linie die Schüler zu Worte kamen.²³

Folgende Hauptaussagen waren in allen Klassen zu vermerken:

Das Schülerurteil fiel insgesamt eindeutig positiv aus, wobei die Äußerungen von „*schon ganz gut*“ bis „*ganz toll*“ reichten. Die Aufgabenstellung wurde als „*ziemlich klar, eindeutig*“ empfunden, das „*Wegfallen der ekelhaften Schreibearbeit*“ und die vorgegebene Ausformulierung sehr begrüßt. Die Dauer erschien hinreichend bis zu großzügig bemessen. Das Anspruchsniveau der Aufgaben werden als „*nicht sehr schwer*“ bis „*noch lösbar*“ empfunden. Einige Kommentare hatten sozusagen didaktischen Charakter:

„*Ich fand die Hörbeispiele ganz gut, denn anhand solcher Sachen kann man mit Sicherheit besser versuchen, ein Ostinato zu erklären oder zu erkennen, als wenn man es in irgendeiner Weise theoretisch auf so einem Zettel macht.*“
– „*Wenn ein Schüler unterscheiden kann zwischen dem Ravel und diesem andren hier, dann kann man wohl sagen, daß er etwas gelernt hat; deshalb sind solche Aufgaben, wo man ein bißchen nachdenken muß, die angemesseneren.*“

Deutlich empfunden wurde, daß nicht nur ein einzelner Bereich erfaßt werden sollte:

„Die Fragen waren breit gestreut.“ –

„Ich glaube, daß der Test total vielseitig war.“ –

„Vom Inhalt her war der Test schon ganz gut, man hat da was von, nicht nur von Beethoven und so.“

Besonders viele Beiträge erfolgten zu den Hörbeispielen; wahrscheinlich deshalb, weil diese innerhalb eines Tests für die Schüler eine neue Erfahrung darstellten. Ihre Verwendung wurde grundsätzlich begrüßt; es gab aber auch Einwände bzw. Änderungsvorschläge.

Kritik wurde auch an der Formulierung einiger Distraktoren geübt.

Auf die Frage, ob sich die Schüler bei diesem Test wohlgefühlt hätten, erfolgte einstimmig eine positive Antwort. Damit ist nicht ausgeschlossen, daß ein Einzelner dennoch Angstgefühle empfand, die ihn nach wie vor davon abhalten, sie zu artikulieren. Aber der großen Mehrheit schien der Test im Gegensatz zu den bisher gewohnten schriftlichen Leistungskontrollen eher Freude als Angst bereitet zu haben.

Die abschließende Frage, was man in Zukunft noch besser machen könne, wurde mit „gar nichts“ oder „genau so lassen“ beantwortet. Man darf aber annehmen, daß bei zunehmender Vertrautheit der Schüler mit derartigen Tests auch die kritischen Einwände zunehmen, was im Sinne einer permanenten Verbesserung durchaus wünschenswert ist.

Über den bisher in großer Zahl herausgestellten Positiva lernzielorientierter Test sollte nicht vergessen werden, daß es auch im Zusammenhang mit diesem Verfahren der Leistungsmessung Probleme bzw. Gefahren gibt.

Als wichtigste erscheinen mir:

- Leicht werden statistisch abgesicherte Testverfahren als perfektionistische Instrumente betrachtet und in den Vordergrund gerückt, obwohl sie im Unterrichtsprozeß an nachgeordneter Stelle stehen müssen, nachdem zunächst die Ziele geklärt, Gegenstände und Methoden ausgewählt wurden. Tests sind lediglich Prüfinstrumente, um die Richtigkeit dieser Auswahlen zu kontrollieren.
- Colwell betont in seinem o. a. Buch: „*A test may be reliable, valid, and usable, but it must also be useful.*“²⁴ Mit anderen Worten: die Brillanz höchster Signifikanzen bleibt wertlos, wenn das Verfahren für die Praxis keinen Nutzen mit sich bringt.
- „*Perfekte Leistungsmeßverfahren bergen die Gefahr in sich, den Lehrer auch in Bezug auf seine Vorentscheidungen selbstsicher zu machen und ihn glauben zu lassen, mangelhafte didaktische Reflexion durch die Qualität der Meßinstrumente irgendwie auszugleichen.*“²⁵
- Der Einsatz von Tests trägt zur Objektivierung der (Zeugnis-)Zensur bei;

er löst jedoch nicht alle Probleme der Leistungsmessung. Viele Lernziele können nicht durch eine Testform erfaßt werden; möglicherweise sind manche wichtigen Lernziele durch überhaupt keine objektive Methode zu überprüfen. Zu klären ist, welcher Stellenwert derartigen Bereichen bei der Zensurenfindung zuzugestehen ist. Zu klären ist weiterhin der Anteil, den die über einen Informellen lernzielorientierten Test erfaßte Leistung an der Zeugnisnote hat.²⁶

Das hier dargestellte Testverfahren vermag nur einen begrenzten Ausschnitt aus dem Unterrichtsgeschehen zu erfassen. Dieser Ausschnitt wird jedoch – korrekte Anwendung vorausgesetzt – mit einer Präzision gemessen, die allen anderen bisher praktizierten Methoden eindeutig überlegen ist.

Zwei zentrale Forderungen ergeben sich für die Musikdidaktik:

1. Leistungsmessung im Musikunterricht muß in ein Evaluationsmodell integriert werden, das den gesamten Prozeß von der Formulierung eines Curriculum bis zur Kontrolle der Kontrollverfahren regelkreisartig umfaßt. Dabei sind gleichzeitig theoretische Rahmenkonzepte als Modelle und empirische Untersuchungen einzelner Bereiche des Evaluationsprogramms (wie etwa der Einsatz von Tests) notwendig.
2. Die Vermittlung von Kenntnissen über Verfahren der Leistungsmessung wird bisher bei der Ausbildung zukünftiger Musiklehrer sträflich vernachlässigt. Es ist ein Konzept zu entwickeln, das darlegt, in welchem Umfang und auf welche Weise die Studenten die Befähigung erhalten, selbständig Informelle Tests und ähnliche Verfahren für den Einsatz in ihrem eigenen Unterricht zu konstruieren; für bereits tätige Lehrer müssen Fortbildungsmöglichkeiten geschaffen werden. Weder standardisierte Tests noch vorgegebene Testreihen innerhalb neuerer Musik-Unterrichtswerke können die Funktion eines lernzielorientierten Tests für den Unterricht des einzelnen Lehrers übernehmen.

Die Realisierung dieser beiden Projekte bedeutet gleichzeitig eine aktive Beteiligung aller Betroffenen vom Hochschullehrer bis zum Schüler. Damit könnte „die Forderung wechselseitiger Begegnung zwischen Wissenschaft und pädagogischer Praxis“, die eine Musikzeitschrift in einem neuen Prospekt als Schwerpunkt ihrer Arbeit bezeichnet, auch realiter eingelöst werden.

Summary

German literature on music teaching offers little help to teachers concerned

with the measuring of school achievements. On the other hand, teacher-constructed tests often lack objectivity, validity and reliability. Therefore in music education a method is needed that is practicable as well as exact. In this context content standard tests are discussed as a way to combine the diagnostic and the didactic functions of a test. The use of three different tests and the results of the analysis of their data are described as an example of how the problem of measuring as part of an evaluative concept can be solved.

Anmerkungen

- 1 Füller, K.: Kompendium Didaktik Musik (München 1977) S. 109.
- 2 Füller, K.: Standardisierte Musiktests (Frankfurt/M. 1974) S. 5.
- 3 Bloom, B. S., u. a.: Taxonomie von Lernzielen im kognitiven Bereich. (Weinheim 1972) S. 240.
- 4 Colwell, R.: The evaluation of music teaching and learning (PrenticeHall 1970).
- 5 Eicke, K.-E.: Die Evaluation des Musikerziehung, in: Forschung in der Musikerziehung 1972 (Mainz 1972).
- 6 Colwell, R.: a.a.O., im Preface.
- 7 Füller, K.: Lernzielklassifikation und Leistungsmessung im Musikunterricht (Weinheim 1974) S. 20 f.
- 8 Ebda. S. 66.
- 9 Diese Systematik ist mehr praxis- als theorieorientiert. Zum einen gehen die Bezeichnungen für das 4. und 5. Verfahren in der Literatur bisweilen ineinander über; zum anderen erscheint Typ 3 in keiner wissenschaftlichen Übersicht (dafür aber um so häufiger in der Schule).
- 10 De la Motte-Haber, H.: Die bildungspolitische Bedeutung des Begriff „musikalische Begabung“, in: Empirische Forschung in der Musikpädagogik (Mainz 1970) S. 13.
- 11 Vergl. Rapp, G.: Messung und Evaluierung von Lernergebnissen in der Schule (Bad Heilbrunn 1975) oder Grubitzsch, S., und Rexilius, G.: Testtheorie – Testpraxis (Reinbek 1978).
- 12 Allgemeine Schulordnung NRW vom 8. November 1978 § 22 (4).
- 13 So Rainer Schmitt in Heft 6/1981 der Zeitschrift „Musik und Bildung“, S. 396.
- 14 Vergl. Plöbl, W.: Lernziele – Lernerfahrungen – Leistungsmessung (Donauwörth 1972).

- 15 Gaude, P., und Teschner, W. -P.: Objektivierete Leistungsmessung in der Schule (Frankfurt/M. 1970) S. 49
- 16 Ebda. S. 89.
- 17 Gronlund, N. E.: Die Anlage von Leistungstests (Frankfurt/M. 1974²) S. 17 ff.
- 18 In einer Podiumsdiskussion zum Thema „Musikwissenschaft und Musik-Lehrerausbildung“ am 2. 10. 1977 in Gießen.
- 19 Eine derartige Untersuchung habe ich vor einigen Jahren anhand des Menuett-Satzes einer Haydn-Sinfonie mit drei Schulklassen durchgeführt.
- 20 Mit der gleichen Begründung lehnen die Herausgeber eines neueren Lehrbuches („Musikunterricht“ Sekundarstufe 1 Band 1) eine derartige Trennung ab. – Vergl. dazu im Lehrerband von Kleinen, G.; u. a. (Mainz 1980) S. 11 f.
- 21 Aus Platzgründen kann dieser Zielkanon hier nicht wiedergegeben werden.
- 22 Z. B. Gronlund, N. E., a.a.O.; Gaude, P., und Teschner, W.-P, a.a.O.; oder Lienert, G. A.: Testaufbau und Testanalyse (Weinheim 1969³).
- 23 Diese Stunden wurden auf Tonband aufgezeichnet, um eine Auswertung aller Schülerbeiträge zu ermöglichen.
- 24 Colwell, R.: a.a.O., S. 41.
- 25 Gaude, P., und Teschner, W.-P.: a.a.O., S. 28.
- 26 Denkbar wäre ein Prozentsatz, der dem zeitlichen Anteil der durch den Test erfaßten Unterrichtsreihe am Schulhalbjahr entspricht; vorausgesetzt, der Test stellt für die Ziele dieser Reihe ein hinreichendes Prüfverfahren dar.

Werner Lohmann
 Scharnhorststr. 12
 D-4230 Wesel 1

Diskussionsbericht

Bei vielen Musikpädagogen ist nicht erst seit Erscheinen der Arbeiten zur „Fragwürdigkeit der Zensurengebung“ ein Unbehagen gegenüber Leistungsbeurteilung im Musikunterricht festzustellen. Es wurzelt in einem tief empfundenen Mißtrauen, die Begegnung des Menschen mit Musik einem möglichst objektiven Beurteilungsverfahren auszusetzen. So wundert nicht, daß auch im Verlaufe dieser Diskussion das besondere Verhältnis des Menschen zur Musik immer wieder durchscheint. In einer ersten Stellungnahme wird registriert, daß sich gerade jene Teilbereiche des Musikunterrichts einer meßbaren Kontrolle entziehen, die zentrales Anliegen des Unterrichts sein sollen, etwa das praktische Musizieren. Es geht nicht um das Heranziehen eines analysierenden Hörertyps, der Höraufgaben spielend bewältigt (Meier). Die Erfahrung wird von vielen geteilt, daß jene Lehrer am häufigsten Leistungskontrollen durchführen, deren Unterricht Qualität vermissen läßt (H. (:h. Schmidt). In diesem Zusammenhang stellt sich die Frage, ob ein „idealer“ Lehrer ganz auf Leistungstests verzichten kann, wenn er seinen Unterricht entsprechend einrichtet. Dagegen wird argumentiert, der Schüler braucht die Beurteilung als Kontrolle seiner Leistungsfähigkeit und als Mittel zur Leistungsmotivation. K. Schmitt empfiehlt mit Hinweis auf einen Beitrag in der Zeitschrift *Musik und Bildung* (6/1981, S. 396) statt Leistungskriterien Verhaltenskategorien so aufzuschlüsseln, daß sie ihren Niederschlag in der Notengebung finden können. Auf verschiedene Ebenen des Verhaltens zielt die Frage nach dem Sinn eines Tests, der das Erkennen eines Ostinato vom Schüler fordert (E. Klausmeier). Ein Test, der nur das Ankreuzen von vorgegebenen Antworten zuläßt, fördert nach Ansicht von H. Gembris die Unfähigkeit, über Musik sprechen zu lernen. H. G. Bastian weist auf das besondere Problem der unterschiedlichen Lernvoraussetzungen und -bedingungen hin. Während ein Schüler die Lösung eines Problems nur unter höchster Anstrengung bewältigt, kann ein anderer dies spielend, vor allem bei entsprechender außerschulischer musikalisch-praktischer Betätigung. Es entspannt sich eine lebhaft diskursive Diskussion, inwieweit Musikunterricht außerschulische Erfahrungen aufnehmen und in die Bewertung einfließen lassen darf. Außerschulische Einflüsse lassen sich nicht verhindern (Böckle), und wie die Erfahrung lehrt, sind Instrumentalisten und Chorsänger die eigentlichen Säulen des Unterrichts (Heinrich). Ein pragmatischer Vorschlag sieht die gesonderte Bewertung instrumental-technischer Leistungen vor (Lohmann). Jede Art der Lernerfolgskontrolle ist mit erheblichem Aufwand verbunden (Gembris). Deshalb ist die Forderung nach einem Vortest, um den

individuellen Lernzuwachs zu prüfen (Tischler) oder die Anregung, die vorherige Lehrerbeurteilung und das erzielte Testergebnis miteinander zu korrelieren (Schmidt-Atzert), wohl unrealistisch. Von Validitätsproblemen möchte H. G. Bastian den Praktiker befreit sehen, und W. Lohmann betont, daß eine Übergewichtung der Testkonstruktion den Blick für die Übereinstimmung von Lernziel, Inhalt und Kontrolle verstellt.

Rudolf-Dieter Kraemer