

0086 d - Title page

0086 d - Title page

ANNOTATED  
INSTRUCTOR'S EDITION

# Wie, bitte?

## INTRODUCTORY GERMAN FOR PROFICIENCY

**William B. Fischer**

*Portland State University*

**Peter N. Richardson**

*Linfield College*



WILEY

JOHN WILEY & SONS

New York Chichester Brisbane Toronto Singapore

O Osted - 0

# Tests and Testing

The Test Bank contains proficiency-oriented or "prochievement" tests for all four modalities, spaced at intervals of about two chapters. The tests are appropriate to *Wie, bitte?* in both their content (selection and sequence of vocabulary and grammar) and the style in which the material is presented. We urge you to use the tests or, if you design your own, at least to maintain the principles of proficiency-oriented teaching and testing. To that end we provide sample tests below. The Test Bank also includes a discussion of testing principles and samples of actual student performance.

Most instructors will likely test less frequently than every two chapters. Since the distinction between teaching and testing should be slight in a proficiency-oriented course, the tests not used as tests could then be used as practice tests. In form they would resemble the real tests, but would lack the psychological pressure. In the latter feature they would resemble the Class Text and Study Text exercises in the corresponding modalities, but would be more formally organized.

## Testing

In a sense *Wie, bitte?* and the entire proficiency "movement" to which it claims kinship were born of a distinct notion of testing, one quite different from the traditional battery of grammar-translation exercises, vocabulary lists, and dictations. The OPI or Oral Proficiency Interview that began evolving several decades ago was an important inspiration for the ACTFL/ETS *Proficiency Guidelines*. They, and early workshops in oral proficiency testing, were the proximate inspiration for *Wie, bitte?* It will not be surprising, then, that we have devoted extra effort to the bank of tests offered here — and that we are aware of at least some of their shortcomings.

Larger issues of testing are addressed or at least mentioned in the *Class Text Instructor's Annotated Edition*. Here we deal with the rationale and mechanics of the *Wie, bitte?* tests in their specific form. Those issues include design, administration, and adaptation. Features specific to individual modalities will be addressed in separate sections.

We take it for granted that teaching for proficiency goes hand in hand with testing and grading for proficiency. Under those circumstances, traditional tests of the kind mentioned just above cannot be the foundation for evaluation of performance, though they may have some limited use. Instead, appropriate tests will be sensitive to modality, will evaluate function and context as well as accuracy, and will emulate genuine communicative tasks. The last of these points can be seen to imply that reading and listening tests will incorporate authentic or at least realistic materials, and that speaking and writing tests will focus on situational tasks. A further desideratum would be that instructional and testing activities should resemble each other, for the sake of that congruence alone, and because both pursuits should emulate real-world communicative activities. A final stricture would be that testing not overburden the teacher, who will be quite challenged for energy if the class is being conducted in a genuinely proficiency-oriented environment.

The *Wie, bitte?* tests reflect these and other pedagogical assumptions, as well as a healthy pragmatism born of the authors' years of experience in teaching (and before that, learning!) introductory German, which include time spent in the Ivy League, at a large Big Ten university, at a moderately selective private university in the South, at a small liberal arts college, and at a non-selective urban institution of mass public higher education. Chief among our axioms, after the commitment to some form of proficiency-oriented testing, were: a) the prime but not monomaniac focus on teaching and testing speaking skill; and, b) the possibility of modifying orthodox proficiency testing methods and standards to fit the environment of the first-year course, among whose most important features are large numbers of students and rapid development of

communication skills. In other words, one needs to be able to test frequently and with considerable discrimination of performance levels, while still maintaining the proficiency-oriented atmosphere and avoiding exhaustion of both students and staff.

Our solution is a bank of tests in each of the four modalities, spaced at intervals of generally two textbook chapters. In their overall nature the tests can be termed "prochievement" tests. We agree with several prominent colleagues that full-dress oral proficiency testing is impractical and unnecessary in lower-level courses, though we will maintain that students taught properly with *Wie, bitte?* stand up well under the pressures of the classic OPI. The *Wie, bitte?* oral tests, then, are limited versions of oral proficiency interviews, with their warmups, level checks, probes, situations, and wind-downs — and, yes, their provision for breakdown and return to level.

Tests in the other three modalities lacked detailed models, in the sense of tests already implemented and carrying the cachet of ILR/FSI/ACTFL/ETS development and use. Nevertheless, the available research and speculation, the proficiency *Guidelines*, and common pedagogical sense as well, offered many resources. The *Wie, bitte?* writing tests focus on equivalents of the oral-proficiency interview "situation," the presentation of a realistic communicative task in considerable detail and yet with sufficient latitude to give students of almost all levels something they can accomplish and something they can struggle with. Many of the writing tests, however, include a sort of "warmup" phase in which lower-level competence is checked. The writing tests, then, are not interactive, as are oral tests; rather, they are "snapshots" of linguistic performance, and therefore subject to the shortcomings of efforts to get everyone "posed" at once.

For the listening and reading tests proficiency-oriented research and discussion furnished more detailed inspiration, though when *Wie, bitte?* was written there were still no generally accepted, practical tests of those modalities in any languages. Nevertheless, certain principles and limitations were clear to us, aside from the general desideratum that authentic materials be used as much as possible. An important inspiration was the collection of articles in *Foreign Language Annals* 17 (October 1984), though other influences include discussions with leaders in proficiency research at the first national workshop in German oral proficiency testing held in Washington, D.C., in early 1983.

Those inquiries and deliberations yielded the result that the basic form of listening and reading tests might well be an objective test whose chief form would be multiple-choice, with simple list-making as an ancillary method of evaluation. To be sure, the tests would not be interactive or adaptive; that is, all students would have to confront the same materials and questions. But — aside from their

positive features — they would have one major virtue: while they would provide a convenient and probably valid enough check of listening and reading, they would not overburden the teacher who was struggling to administer oral tests and provide individual evaluation of writing tests. To put it quite succinctly: the teacher who undertakes systematic individual oral testing, and who likely also wants to devote some time to individual evaluation of writing skills, needs to be able to test the other modalities with dispatch.

### TESTS — Administration

The Test Bank offers tests in each of the four modalities at intervals of every other chapter. To avoid bunching of tests, the intervals are staggered. Oral and writing tests fall on odd chapter numbers, listening and reading tests on even numbers. Thus in a quarter during which 10 chapters were covered, 20 tests would be available. We doubt that such frequent testing is necessary or wise. Instead, we have provided such a great number of tests so that instructors may fit the *Wie, bitte?* package to varying academic schedules, will have alternate tests in order to maintain test security, and can use some of the tests for practice, whether in class or in lab or home study. For additional test security the supplementary materials include a set of print realia (in this volume) and a separate cassette. Most of the listening and reading tests can be administered with machine-scorable test forms, such as SCAN-TRON™.

Reading tests use side 1 (items #1- ); listening tests use side 2 (items #51- ). The tests are laid out for economic copying from the page; they are also available on computer diskettes as word-processor files, should you wish to customize format or content, or to print mimeograph stencils.

The listening, reading and writing tests are each calculated to require 20-30 minutes, at least after those for the earliest chapters. Except for the listening tests, which require careful performance of aural material or use of a tape player, the tests are no more difficult to administer than conventional tests. We recommend three tests in each modality per quarter, three or four in a semester. Thus per quarter between four and five hours of class time would be devoted to testing — about 10% of the time available in a class that meets four periods a week. That amount differs little from what conventional testing usually requires.

We would point out too that the tests resemble the various *Wie, bitte?* exercises in the several modalities, and that both exercises and tests emphasize real-world tasks and the use of common sense and general intelligence; so the tests can be regarded as learning experiences as well. But because the tests differ from those commonly anticipated by beginning language students, the instructor may want to make the first test count significantly less than subsequent ones, or may even want to

administer the first available test in each modality as a non-counting but seriously evaluated practice test. Thus there can be “dry-run” tests in reading and listening for Preliminary Chapter 2, and in writing for chapter 1. Not many programs will be able to afford a non-counting individual oral test, though we heartily if wistfully recommend it.

Scoring, evaluation and grading pose several difficulties, with additional special problems for the individual modalities. The most general issue is how to weight the several modalities. Two factors, somewhat contradictory in their implications, deserve attention. In their statements of desiderata many or even most learners prize the practical ability to speak the language. At the same time many learners tend to believe, by conditioned cultural reflex, that written tests are the “real” tests, and that tests of analytic or intellectual knowledge (i.e., grammatical principles) are the real gauge of performance in foreign language courses.

We suggest that you carefully prepare your students to regard listening and oral tests as important; you may even wish to assign the greatest weight to oral tests, if you can administer them frequently and carefully enough. At Portland State University, where every student has three ten-minute individual oral tests outside class each quarter, the grade is apportioned as follows: 30% speaking, 25% listening, 25% reading, 20% writing. Students who wonder why three ten-minute oral tests should count more than the three longer tests in each of the other modalities are reminded that opera performers are paid not for the relatively few minutes they spend on stage, but rather for the lengthy preparation that precedes the moments in the spotlight. Probably you will want to weight tests taken later in the grading period more heavily than those at the start, especially in the first part of the course.

Pegging test results to some more enduring standard than a curve for the current test is also problematic — as it is with conventional tests. While *Wie, bitte?* was designed with the ACTFL/ETS *Proficiency Guidelines* ever in mind, and while those standards can be used to describe progress throughout the course, one cannot simply declare a correlation between the *Guidelines* and performance on a particular test, or — much worse — facilely mandate correspondences between grades and proficiency levels. The *Wie, bitte?* tests are not in themselves proficiency tests, and the intervals between proficiency levels cannot easily be mapped onto an ABCDF scale or, much less, equated one for one to equally divided intervals of seat-time. The matter is vexed still further by differences among institutions, in both their general rigorousness and in the number of hours per week their language courses meet.

Nevertheless, for certain of the tests we have offered descriptive standards of performance, which are clearly based on notions of proficiency. And at our own institutions we do assign grades and have

some notion of how they correspond to levels of proficiency. We have never made an ironclad demand that students attain a certain proficiency in one given modality in order to receive this or that grade, such as Intermediate-High in speaking for an A. But we do use the *Guidelines* to gauge our students' progress and to ensure that standards are maintained over time, without creeping grade inflation or deflation. With regard to the "production" modalities of speaking and writing, it is our conviction that, at our own institutions, the top-notch student who came to the course as a *bona fide* beginner will likely perform at the Intermediate-High level in one of those modalities, and at the Intermediate-Mid level in the other. The average student will likely perform at Novice-High in one of those modalities, and somewhere in the Intermediate area in the other. We remind our readers that there have been as yet no studies of proficiency levels attained by students in proficiency-oriented first-year German courses using textbook packages designed expressly to further proficiency.

Still larger issues must be considered, and notes of caution and encouragement should be sounded. Proficiency-oriented tests, and perhaps especially writing tests, may occasion frustration or even alarm in the instructor. One may feel that so much has been taught and so little learned, even by the better students. One may also compare ruefully the performance of classes on conventional grammar-translation tests to those of students on *Wie, bitte?* tests. The alarm is probably unwarranted, and the comparison is invalid. Conventional tests ("Now replace the noun phrase with a pronoun in the appropriate case.") make their targets quite evident, either explicitly or — to the knowing student — implicitly. The student may then concentrate on the recollection and production of the target material.

To put the matter more favorably: proficiency-oriented instruction and testing may provide valuable diagnostic tools that reveal genuine problems that conventional testing may fail to identify. The *Sie / sie* 'you / she' distinction provides an egregious example. Teachers whose chief focus is analytic grammar, especially as demonstrated on conventional written tests, may not suspect that their students may well lack genuine proficiency in making that distinction, either in comprehension or production, because the exercise or test "telegraphs" or gives away its target; the verb chart or similar transformation setup tells the student which categories to anticipate. A simple experiment will test the point: In a plausible context, mix utterances (or probes for utterances) that focus alternately on the second and third persons, with reasonable if somewhat unexpected transitions, and then await the outcome ("So. Ihre Schwester heißt Linda. Und was studieren Sie?" "Sie studiert Musik."). Many students who seem to exhibit considerable analytical mastery of, for example, the pronoun system, will still fail to demonstrate proficiency in it, either in comprehension or in

spoken or written production. The same will likely prove true of tense distinctions, even in second- or third-year students.

Proficiency-oriented teaching and testing, like life, pose problems to be solved individually and creatively with the full range of means at hand, and it is recognized that practical skills are acquired gradually and employed as they are felt appropriate. Those who seek a common sports analogy might contemplate the difference in baseball between the inflated achievements of batting practice, where the pitcher and perhaps even the pitch are known, and the hard-fought struggle of the real game, where one must deal on the spot with any eventuality.

### TESTS — Adaptation

Some instructors may want to adjust the content of the *Wie, bitte?* tests, whether to strengthen test security, to decrease or increase the length of a given test, or to fit a test to, say, the preceding or following chapter. Altering the listening and reading tests will often be a trivial matter of substituting single words or short phrases, either within the given chapter or with items from another. Frequently even entire tape segments, for example the time announcements or weather reports, may be substituted for segments of a similar kind. Entire sections may be identified as relevant to a chapter not yet presented, and therefore deleted or, by direction to the students, ignored; generally such sections are toward the end of each test.

Since the tests are proficiency-oriented and therefore examine broader competence rather than merely discrete-point knowledge, less adjustment may be necessary that would seem requisite initially. The oral and writing tests are particularly amenable to use for chapters later than their original specification; one simply raises the standards of accuracy and looks for the production of material that has indeed been presented. Thus the writing test for chapter 3 involves a postcard note to the proprietors of a hotel in which the student has stayed. At that point the student has only *sein* in the past tense, and is therefore limited in describing what has happened elsewhere on the journey. But the same narrative task can be posed to advantage later, either after chapter 9 (preterit of modal verbs), or certainly after the present perfect has been introduced (chapters 12-16, 22) and the preterit system extended to ordinary verbs of high frequency (chapters 21, 24). The same principles apply to the oral tests. To be sure, many tests could be used in second- or third-year courses as well.

Some instructors may wish to make major alterations in the *Wie, bitte?* testing plan while still maintaining the proficiency orientation. If the multiple-choice format of the listening and reading tests is considered too mechanical or confining, a ready alternative would be tests that resemble the Study Text exercises in the corresponding modalities; but difficulty in evaluating performance

objectively and conveniently should be anticipated. Some teachers, too, may want to attempt integrated or multi-modality testing, where — for example — the student might first read or listen to something and then undertake a related speaking or writing task. Such tests could convey very pointedly the complexity and realism of human communication; but they, too, might involve problems of evaluation, since it might be difficult to decide in which modality weakness in communication might have its source.

We suspect, though, that it will be in oral testing that the greatest changes will seem necessary. Logistical difficulties or sheer lack of time may preclude regular individualized oral testing outside class. Several alternatives, other than simply having no oral evaluation, have been proposed in the professional literature.<sup>1</sup> Independently taped tests do not seem to constitute a first choice, since they lack the element of communicative give-and-take. More to be recommended are in-class evaluations, either frequent but brief measurements of individual performance, or lengthier assessments of group work. We would remark lastly that many courses include one hour in the language lab each week. Our own experience suggests that a ten-minute individual test every few weeks has a very positive effect on learning patterns and will be remembered in detail by the student. That advantage may make it worthwhile to substitute oral testing sessions for at least some conventional lab work.

#### ORAL TESTS

*Wie, bitte?* is not "just" an oral proficiency book, as a glance at the Study Text will show. Nevertheless, the development and evaluation of speaking skill have been prominent parts of the notion of proficiency. For that reason, and because an active and interactive classroom is a happy and — more important — a productive classroom, the Class Text focuses on oral proficiency. It would be difficult to teach with the *Wie, bitte?* program without somehow evaluating progress in speaking. Students are not stupid; whatever we preach, they

<sup>1</sup>Brown, James. "RSVP: Classroom Oral Interview Procedure." *Foreign Language Annals* 18 (1985): 481-86. Duncan, Cynthia K. "Evaluating Speaking Skills in the Classroom: A Workshop for Teachers." *Foreign Language Annals* 20 (1987): 15-23. Kaplan, Isabelle. "Oral Proficiency Testing and the Language Curriculum: Two Experiments in Curricular Design for Conversation Courses." *Foreign Language Annals* 17 (1984): 491-97. Larson, Jerry. "Testing speaking Ability in the Classroom: The Semi-direct Alternative." *Foreign Language Annals* 17 (1984): 499-507. Pino, Barbara Gonzalez. "Testing Second Language Speaking: Practical Approaches to Oral Testing in Large Classes." *Northeast Conference Newsletter* 24 (Sept. 1988): 14-16.

will tend to take their cue from our tests and our grading, especially if we then offer conventional written examinations of analytic grammar, vocabulary list memorization, and so on. But students are also conservative. We may preach and teach for proficiency, and in particular for oral proficiency, and our students may consciously assent to that goal. Still, because of years of social conditioning, many students will share the general public image that, at least in the academic setting, the "real" test is a written examination of analytic grammar, vocabulary list memorization, and so on. The *Wie, bitte?* oral tests attempt to dispel that notion, and we are convinced that some such oral testing, with corresponding grading, is vital to a proficiency-oriented program. Here we must regard that conviction as axiomatic and proceed to a discussion of means.

The teacher who has decided to institute oral testing must then decide how, and how often, to do it. Logistics suggests the limit at one end of the spectrum, while common pedagogical sense indicates the boundary at the other. Oral tests must be administered individually or, at worst, to small groups of students, and each test must be long enough to permit a fair attempt at eliciting a valid sample of speech and to allow for courteous human interaction — certainly a matter of several minutes at least. Conducting a term's only oral test during the final exam period would be extremely unwise, and for several reasons. All of us have bad (and good) days, so obtaining a representative sample would be unlikely — and the additional stress of having but a single test to prove one's ability would make the student doubly anxious — unless the test were known to count very little, which would nullify its value. The message that oral proficiency was really important would lack immediacy during the course, and the student might wrongly assume that cramming would suffice for that odd and, in all likelihood, hitherto never encountered test of speaking. Lastly, rare indeed would be the instructor who could afford to devote hours of time to a marathon of meticulous and momentous oral testing during finals week.

On the other hand, extremely frequent oral testing — one test each week, say — would be exhausting and unrewarding. Even during the early weeks of a first-year course, progress is not so momentous that it need be measured every few days. Oral tests administered early in the course have the benefit of establishing trenchantly the importance of speaking skills, but a chief disadvantage is their awkwardness; students are quite nervous, and they have virtually no room to maneuver linguistically. We therefore recommend oral testing at intervals of every few weeks — a minimum of three times a quarter, or three or four times a semester.

Here follows a description of the oral testing program that has been conducted for a decade now at Portland State University, a non-selective urban

institution with many commuting students and many having significant work and family obligations. The class meets four hours a week, with 25 or more students per section at the start of the course. In a recent quarter two sections were taught by a tenured faculty member (one of the authors), two by new teaching assistants in the departmental master's program, and one (at night) by an experienced adjunct faculty member holding a master's degree.

Early in the course we conduct an orientation in proficiency goals and testing mechanics. Staff members coordinate their schedules to allow a list of testing hours sufficient to accommodate, at a hypothetical rate of five students per hour, the entire enrollment sometime during a two-week testing period, with some flexibility. Thus the 150 students expected for the first oral test might require about 25 hours of testing time over the two week cycle. Each section instructor, then, is responsible for 2-3 hours per week of testing, assuming that instructors new to the program will not need to be accompanied by experienced colleagues.

During the week before testing begins, sign-up lists are circulated in class. They are accompanied by individual reminder slips, which carry a statement about testing goals and policies and offer room for the students to record their scheduled testing time. The staff meets to discuss testing format and standards. The tests are conducted, with time perhaps for face-to-face diagnosis and advice. Lost students are tracked down. Testing standards are tightened up and new material introduced as the test cycle continues. Tests in other modalities may be conducted in class during the same period, though more of them are scheduled during the weeks when there are no oral tests.

Pedagogical expertise is of course important in the administration of oral tests, though we hope that the *Wie, bitte?* test scripts will be helpful to even relatively inexperienced testers. Not to be neglected, however, is the bureaucratic aspect. Large numbers of students must be encouraged to adjust to new patterns of behavior, and there will always be difficult cases. Scheduling must be orderly, and so must record-keeping. The latter involves not merely reliable storage of overall results, although that is indeed vital, since students who have put forth their best efforts during a strenuous oral test will want them to count. It is best to maintain as well a system of detailed note-taking during the tests, so that evaluations can be substantiated and reviewed. Samples of performance profiles and evaluation forms accompany the first few test scripts. Large programs will likely need systematically organized ways to preserve and check records, perhaps with a computer data base. Post-test review of performance, perhaps in class, will also require some effort.

The prospect may seem daunting, but the rewards — in both pedagogical validity and student performance — can be great. The instructor should remember that such testing can replace to at least a

significant extent the tedious and often evidently pointless business of checking and grading conventional homework and tests. Similarly, tests in other modalities have been designed to save time and effort in grading. A last and more subjective benefit should be mentioned: students in courses that emphasize face-to-face communication often come to value the element of human interaction they may well miss elsewhere.

The scripts for each oral test describe the linguistic targets, outline major sections of the interview, and provide typical question items, with either models in brief German or topic summaries in English. Oral proficiency tests are interactive and adaptive; the test evolves to fit the circumstances. *No two tests will be the same, and you should not attempt to present every item to each student.* The basic script is aimed at the student who is progressing comfortably but not superbly. Provision is made for alternate formulations of single items, and for omission or alternation of entire sections. The overall test follows the pattern of an oral proficiency interview: warm-up (greetings, brief everyday matters); level checks and probes, including a role-play situation and earnest attempts to detect the student's breakdown level; and wind-down (thanks, small talk, farewell). The role-play situations, if they are brief and can be presented simply, are incorporated within the outline; more elaborate ones are printed separately, like the well-known OPI cards, and you may well want to present the text to the student, allow a minute for perusal, and then proceed, with the student consulting the text if necessary. Each student should be thoroughly tested, but there is no point in beating a dead horse. With below-average students you may well need to omit some of the more difficult sections. Even a role-play situation may be inadvisable, at least early in the course, though the situations provided for most tests offer a range of difficulty and theme. For test security you will want to vary the test items anyway, and if you conduct your tests over a period of more than a few days you should update the material and raise standards of performance.

#### TECHNIQUE

The Bibliography lists several discussions of oral testing technique, most notably an article by one of the *Wie, bitte?* authors (Fischer). Also useful, if you can obtain a copy, is the handbook used in ACTFL oral proficiency testing workshops.

Oral tests must be conducted in a relaxed manner. They should resemble not interrogations but rather conversations, with the examiner as a somewhat more insistent version of the notorious friendly, loquacious representative of the target-language culture. Transitions between the topics or linguistic features being checked should be plausible, and the examiner should always have ready fallback

reformulations or changes of subject should serious obstacles be encountered.

Nevertheless, the tests must be thorough and — given the time restrictions most instructors face — efficient. Two precepts are especially important, but both may run counter to the personality of the typical teacher, who wants so much to teach ever more material and to view success. Here, too, less is often more — or at least better. The examiner should present enough checks and probes to obtain a valid profile of performance, of course, but should not attempt laboriously to lead the student through items which are manifestly very difficult or, much less, hopeless. Secondly, the teacher should leave most of the speaking to the student, by presenting items briefly and in simple language, by learning probe techniques that encourage copious speech production, and by developing extra patience in waiting for replies. To be avoided are Yes/No questions (except in early units or with poorer students), phrasing which reveals target structures and vocabulary, and — above all — speech which is artificially slow and grossly exaggerated in intonation. Here are some useful generic tactics:

- 1) blatant pauses, perhaps with a raised eyebrow or a brief interjection ("Oh?")
- 2) leading phrases ("Und [dann]?" "Allein?"), suppositions ("Ich glaube, Sie studieren Biologie."), and W-questions ("Wann?" "Wie?")
- 3) formulations which virtually demand sentence creation, for example double questions ("Wie alt sind Sie, bitte, und was studieren Sie?") or fragmentary follow-ups ("Was machen Sie heute abend?" . . . [student answers] . . . "Und am Wochenende?")

It is very important that you keep a careful record of the test performance, either by taping it or, more likely, by unobtrusive note-taking during the test. (Inexperienced testers may want to double-staff their tests: one person administers the test while the other makes notes.) For several of the early tests we offer model note-taking forms with provision for three types of evaluation (major section, single feature, global feature) and for a grade assessment. If time permits, you may want to discuss the student's performance with him or her right after the test, offering praise for good points and advice about improvement. Our experience suggests that most students appreciate immediate assessment of their performance; and almost never is an informed grade disputed, for the student will have a distinct memory of many parts of the test.

For tests in the early chapters we also provide descriptive profiles of performance so that performance may be judged by comparison to absolute standards. The profiles, like the ACTFL/ETS *Proficiency Guidelines*, generally give first an assessment of function, both because it is useful to estimate overall performance before examining specific features, and because functional ability should be given more importance than individual forms. The highest category of

performance does not represent errorless mastery of all material previously presented, but rather the proficiency that — in our experience — will likely be exhibited by excellent learners who have not previously studied the language. The point bears amplification: *Wie, bitte?* presents material whose absolute mastery would constitute proficiency at the ACTFL/ETS "Superior" level in speaking; but only very exceptional students will attain "Advanced" proficiency, and indeed the most that might ordinarily be hoped for from the best students — perhaps the top 10% at typical institutions — is "Intermediate-High."

Care should be exercised in correlating test performance to proficiency levels and in equating the latter to academic grades. At least in the early part of a course based on *Wie, bitte?*, performance that seems to qualify for a certain level of proficiency may not represent true proficiency. Although *Wie, bitte?* follows the *Guidelines* closely, not even it could present all of, say, the Novice-High elements early on; thus a significantly different level of proficiency might be assessed if the interviewer were to select a different one of the contexts that might legitimately be checked in Novice-High interviews. But an Oral Proficiency Interview conducted with at least reasonably capable students in the latter part of the program would likely reveal their genuine proficiency levels, since by that time the appropriate contexts and content areas have been presented. That is, OPI assessments of the performance (or lack of performance) of *Wie, bitte?* students at the levels of Intermediate-Low or higher would likely be valid.

While the various profiles of performance can indeed be equated by instructor fiat with academic grades, an attempt to define grades by proficiency levels, perhaps for purposes of planning curriculum, should be undertaken only with extreme caution. While ABCDF grades usually reflect some sort of bell-curve distribution, with the better grades within the range not only of the bright student but also the industrious if not supremely intelligent one, the sequence of proficiency levels resembles an ever-steepening incline in terms of functional ability and the time and effort needed to acquire it. While the incline is fairly moderate at least up to the Novice-High level, it becomes particularly steep between Intermediate-Mid and Intermediate-High. Thus while Novice-High or even Intermediate-Low oral proficiency may be within the range of more than a few students by the end of the first semester or even academic quarter of the course, and Intermediate-Low or even Intermediate-Mid may be quite possible by the end of the second quarter, one simply cannot expect advancement to Intermediate-High by the end of the year.

It would be utterly wrong, then, to declare that an A will be awarded for Intermediate-Low at the end of the first quarter, for Intermediate-Mid at the end of the second, and for Intermediate-High at the end of the year. Similarly, one cannot assign grades synchronically, correlating for example



Intermediate-Low with A, Novice-High with B, and so on. The issue is complicated by the need to consider the other modalities, either coequally or in weighted form, and with or without provision for cross-compensation. Therefore instructors will still have to determine their own standards and calibrate them finely. But the tools of proficiency evaluation can permit one to ascertain that teaching and testing complement each other, and to maintain constant levels over time and among classes that may vary in student quality.

### LISTENING TESTS

*Wie, bitte?* attempts to reinforce and measure skill in listening, a modality that is often neglected, whether because it is regarded as inferior to exercise in reading, writing, and analytic grammar, because it is taken for granted as a transparent or passive ability that needs no attention, or because the instructor is unsure how to teach and evaluate listening skill. Correspondingly, many programs lack genuine listening comprehension exercises; most lab tapes use listening as a means to exercise other skills such as speaking or grammar transformation.

The *Wie, bitte?* listening tests represent a compromise between full-scale proficiency testing and the demand for efficiency in the face of the instructor's ordinary time limitations and the desirability of conducting serious oral tests. The tests are not interactive and adaptive, but they are contextualized and they do employ authentic language. They are also function- and task-oriented, in the sense that the items require the student to comprehend aural information of hypothetical real-world relevance. The tests do not demand overt application of analytic grammar knowledge, although they can serve to evaluate such competence. Several tests, for example, ask the student to decide whether the action in an utterance is finished, is going on, or has not yet begun. The directions do not mention the grammatical categories of past tense (or present perfect), present tense, or future tense. Thus the student does not get caught up in technical terminology, and the testing can be undertaken in terms of function rather than grammatical categories. One notes that the distinctions just described need not be tied to single tenses, but may be used instead to check comprehension of time phrases and other parts of speech besides verbs. The language of test administration is English, so that a listening test does not turn into a test of the ability to read tests.

Overall test administration is straightforward. Most items on most tests are multiple-choice, so that the tests can be graded efficiently, perhaps even by machine. Some tests require hand-correction of a few short note-taking items. Occasionally a more elaborate procedure is used: the student takes notes or marks choices after one listening of a passage, turns those answers in, and then answers other

questions after a second listening. The purpose will be evident: the first listening tests global comprehension, after which more detailed comprehension is tested without the need awkwardly to conceal global information.

Each listening test has a script with directions about test administration and items to be performed. Many tests also use taped materials. The items which are performed "live" should be read in a clear but not exaggerated voice at natural pace. Items which involve low-level language and simple categories of choice are read only once; more complex items are read twice.

While whatever is the same for all students is fair in a certain sense, the taped portions of the test should be played on the best available equipment. Small monaural portables are generally unsatisfactory for any but very small classes. One should remember that what appears acoustically clear to the teacher, who benefits from psycholinguistic factors, may not be at all clear to the student. For classes of up to several dozen students a "boom box" should suffice; a tape counter will be useful.

One should not underestimate the psychological stress posed by listening tests, particularly those which employ authentic materials and which may expose the student to a veritable stream of language while asking only for low-level comprehension. Students are accustomed to tests where they read and write, where no words appear that are not in the book, and where re-reading can aid comprehension. Many also dislike the risk-taking and drawing of inferences that are so important in proficiency-oriented instruction. If they aren't sure they have understood everything, they think they can't understand anything. There are several ways to reduce anxiety: in-class review of Study Text listening exercises, discussion and demonstration of listening strategies, and "dry-run" exercises with Test Bank materials that will not be needed for actual testing (e.g., test three times in the first quarter, at the end of chapter P-2, 4 and 8, and use the tests for chapters 2 and 6 for practice). The instructor should also prepare for the test by rehearsing the script and double-checking tapes and equipment so that the material is presented correctly and smoothly. We recommend that the directions and examples for the test sections be read aloud as the test is administered. Be sure the students fully understand the directions for each test section before they begin it.

### READING TESTS

Of the four kinds of tests in the Test Bank, the reading tests are the simplest to administer. Most items are multiple-choice and therefore machine-gradable. The only special need will be for good reproduction of the reading texts themselves, and especially of the realia used in most of the tests. Note

that there are some sections that check the ability to recognize correct spelling, punctuation, and so on. Be sure that the reproduction clearly permits distinction of significant differences between characters and symbols (a/ä, l/!), especially where German and English conventions differ (e.g., noun capitalization, *ich* vs. *Ich* in mid-sentence, comma vs. period in prices, period vs. colon in clock time). Where budget allows, sets of photocopies of realia can be prepared and then reused. Other instructors may wish to use overhead transparencies. In a pinch some of the authentic material could be retyped and mimeographed, although that practice would seem scarcely worth the effort and would deprive the student of the typographical and design cues that are important clues for reading comprehension in real-world situations.

Realia for the reading tests is located in the Class Text, the Study Text, and — for extra security — in the Test Bank. While word of mouth may tip off new students that a given test uses realia that are readily available, the compromise need not be fatal. The Test Bank contains sufficient material that tests can easily be varied from term to term; items that do not use realia can be redesigned even more easily. And should students absorb the general message that the material in the books will be on the tests, so much the better.

While student anxiety is not as great with proficiency-oriented reading tests as it is with listening tests, some caution and preparation are advised. The spoken word may leave little trace on the mind, but print does. Thus some students may object that "We haven't had this word yet," or "But this word isn't in the book." Similarly, some students will complain that the texts or the tests are too long. Defensively it can be noted that in no case does successful management of a test item depend on comprehension of language that has not yet been systematically presented. One may point out as well that on multiple-choice tests, unless there is a penalty for guessing, random responses will produce some "correct" answers; hence the need for a large number of items, and for some quite difficult ones. But the best defense is a good offense, and preventing is better than curing. In study outside class, and in classroom review, students should acquire tolerance for unfamiliar language and should learn the techniques of skimming, scanning and inference. That is to say, they should learn (or recall) how to read. We are convinced that many students possess adequate functional command of the strategies of reading in the real world, but that they abandon — or are robbed of — those strategies at the classroom door when they undertake reading in an academic environment, especially in the traditional foreign-language classroom.

The rationale for English as the language of test administration is presented in the discussion of listening tests.

## WRITING TESTS

The *Wie, bitte?* writing tests are contextualized, situational tasks that nevertheless target well-defined ranges of structures and vocabulary. They can be viewed as written equivalents of oral proficiency interview situations, but with allowance for the difference in modality. Above all, the writing tests are not like conventional written tests, which generally consist of grammar-oriented transformation exercises, checks of vocabulary lists, and perhaps a dictation section.

For some chapters the writing test offers shorter initial sections that check handwriting, contextualized vocabulary, or similar relatively discrete features. It is in keeping with the proficiency orientation of *Wie, bitte?* that handwriting be checked early for general adherence to cultural norms, since proficiency means the ability to communicate within the culture, and genuine communication is sensitive to modality. For similar reasons vocabulary testing is not a matter of discrete-point matching, but rather of contextual association; one tests the student's ability to produce a nexus of lexical items likely to be encountered or needed in association with each other, and one emphasizes and therefore hopes to reinforce the faculty of association rather than mere list memorization.

But the heart of every writing test, even the very first one, is a situation, the description of a communication that might plausibly have to be delivered in written form in a German-speaking culture. The tasks are mature ones that might be undertaken by adult native speakers, but even so they can be accomplished by someone proficient at the current level of instruction. In other words, they are designed to be appropriate in function and content/context. Thus there are no non-contextual "essay" questions or erudite translation exercises, although we note that for many such traditional exercises one can supply a plausible situational context. The presence of context is vital, just as in listening and reading the presentation of language in its original setting and form is important, and for the same reason: genuine discourse is always tied to a context and a purpose, and those elements in turn affect linguistic form at all levels, from style or suprasegmental features down to choice of lexicon.

Those considerations dictate the formulation of the writing test situations. Two features will be noted immediately: the situations are posed in English and they seem verbose. English is used for two reasons: the writing tests should not be reading tests, and one does not want to provide the student with the target structures and vocabulary. The verbosity is actually a carefully calculated strategy with three justifications: since the tests are contextual, but only fictively so, the context must be made abundantly evident so the student will catch the spirit and feel free to imagine and to communicate abundantly; the copious detail contains many probes for production of specific

features recently introduced in the text; and the phrasing is carefully made indirect and even colloquial, so that the student cannot hope to translate, and yet the task is couched in a fairly simple though not colorless idiom that aims to be quite comprehensible to students who, though their native language be English, may not be masters of academic prose.

Administering the tests is a fairly straightforward matter. You may want to provide scratch paper for note-taking and a first draft. Time limits should be enforced fairly, of course, but beyond that lies the problem of what time should be allowed. For most tests we advise 20-30 minutes, since in most cases test time must be taken from class time and since the student who can do the task at all can likely do it in that amount of time. But since the writing tests are communicative tasks rather than discrete-point examinations, some students may initially find them disturbing. They may assume — wrongly! — that a conservative approach will yield a higher grade because it involves less chance of error; a sermon about proficiency levels and the benefits of risk-taking may be in order. They may also wonder whether they have truly "finished" the task, and in the "right" way. Given time, some students would still be writing much later, and yet would not produce a linguistic sample that would rate significantly higher than what a much shorter time would yield.

Viewing the matter from another perspective, one may remark that the person who needs several hours to compose a simple telephone message is not functionally competent, since in the real world such messages must be managed in far less time. But in the real world, of course, tasks are posed by genuine circumstances, not by elaborate written descriptions. Thus we have gauged the timing of the *Wie, bitte?* writing tasks as follows: enough time to read the task at a comfortable speed, and then twice as long as it would take to carry out the task in one's native language, plus a few minutes. A post card bringing an acquaintance up to date on one's travels over the past few days should not take more than about five minutes to write in the native language. Such a task on a writing test, then, might involve fifteen minutes — provided the task does not ask the student to do the impossible with currently available resources.

Since the writing tests are not discrete-point tests, they cannot be graded with simple answer keys. One must ask, instead, what functions the student has carried out, how well, and with what means. Such grading — as was discussed in the section on oral testing — requires global assessment of function and then an examination of detail, with allowance for compensatory skills and alternate formulations. But the instructor must nevertheless be able to grade the tests in a reasonable time.

We recommend the following procedure. The test materials include profiles of performance for early tests; they can serve as models for profiles for

other tests. Note that the profiles are not checklists that enumerate everything that must be demonstrated, but rather descriptions of typical performances. After the profiles have been studied, the grader should read each test rapidly, in a minute or so, and then record a concealed grade, without otherwise marking the tests. A second or even a third reading, perhaps in somewhat greater detail, can yield additional assessments. The technique is particularly useful in large courses with several instructors, who can then check interrater reliability and help maintain standards. The final reading, conducted by the main instructor if there are several readers, is followed by comparison to the earlier assessments. A satisfactory level of interrater reliability can be attained with but a little practice — well within the range of graduate teaching assistants supervised in conscientious pedagogical programs. Grades may then be assigned and weighted in accord with the principles and caveats advanced in the discussion of oral testing.

What one does beyond that is open to deliberation. Some instructors may want to regard a test as simply a test, and treat the matter as closed. Most, however, will want to provide their students some sort of feedback beyond a terse grade. Here the issue involves, as usual, problems of available time and pedagogical strategy — for which there may be a happy compromise. The instructor who completely rewrites every test, and especially the poorer ones, into stellar German will spend a great deal of time that may well be poorly required, either with stultified indifference or with the honorable if not outstanding student's frustration that so much was attempted and seemingly so little attained, with little to learn from it. Our advice instead is to exercise conservative error correction. Point out examples of significant errors, either by correcting and explaining them in detail or by simply flagging them as errors, and limit marking and correcting to what can genuinely help the student. A rule of thumb might be that one draws the student's attention to errors that, if corrected, would advance the grade by one level, say from C to B. Thus the student who, in chapter 5, is still virtually oblivious to the notion of conjugation should not be overwhelmed with reminders about the stem-vowel changes that were presented in chapter 4. On the other hand, the student who is consistently applying the standard principles of regular verb conjugation, even to modal verbs, might benefit from such reminders.

A related matter is the appreciation of error patterns that might actually represent positive learning and therefore deserve some credit. The phrases *\*zum Konditorei*, or even *\*zur die Konditorei*, though incorrect, reveal inchoate competence in the dative case, and some notion of gender. Encountered in tests for the early chapters, they should receive some appreciation, in comparison to *\*nach das Konditorie*, though of course *zur Konditorei* deserves higher approbation. In later chapters, of course, that specific error should

not be regarded so benevolently; by then there will be other such significant, fruitful error patterns, such as \*Ich habe gegeht in chapter 12.

Some instructors may choose to pursue conservative error correction in yet other ways. The presence of errors might be noted by simple marks, with the proviso that the student, to earn some increase in grade, rewrite the test with a certain measure of improvement. The instructor may also prepare excerpts of significant responses, including examples of excellence as well as error, which can then serve as the focus of brief discussion and further exercise.

---

We would like to thank our adjunct faculty, teaching assistants, and practicum students who contributed test materials to this program.