technische universität
dortmund

Willkommen Welcome Bienvenue Bienvenido Benvenuto

Assessing Student Learning Outcomes

Teacher Perspectives on Assessment Formats

Assessment Guide for English

zhb. Zentrum für HochschulBildung
Bereich Fremdsprachen

Assessment Guide for English

Assessing Student Learning Outcomes: Teacher Perspectives on Assessment Formats

Edited by Meni Syrou & Helen Horner

Table of Contents

Einleitung

Meni Syrou

In der zunehmend globalisierten Berufswelt sind fundierte Fremdsprachenkenntnisse für Hochschulabsolvent*innen unumgänglich; die kompetente Beherrschung von mehreren Fremdsprachen gehört zum akademischen Berufsprofil. Die in der Bologna-Erklärung 1999 dargelegten Eckpunkte *Mobilität*, *Wettbewerbsfähigkeit* und *Beschäftigungsfähigkeit* für Studierende und Hochschulabsolvent*innen sind ohne fundierte Fremdsprachenkenntnisse nicht zu realisieren. Universitäre Sprachlehreinrichtungen besitzen somit die Aufgabe, eine adressatengerechte, das heißt den spezifischen Bedürfnissen von Studierenden Rechnung tragende Fremdsprachenausbildung zu gewährleisten. Fremdsprachenkenntnisse stellen nicht bloß eine Schlüsselqualifikation dar, sie beschreiben vielmehr ein eigenständiges Kompetenzprofil, das eng mit der Fachlehre verknüpft ist. Fremdsprachkenntnisse auf einem akademischen Niveau bilden die Grundlage, auf der die Vermittlung von wissenschaftlichen Kenntnissen im international geprägten Hochschulkontext überhaupt möglich wird (vgl. Vogel, 2009, S. 12).

Als universitäre Sprachlehreinrichtung fokussiert das zhb Bereich Fremdsprachen die Vermittlung von Sprachkenntnissen in hochschulspezifischen und berufsorientieren Kontexten. Eine vorrangige Zielsetzung unseres Bereichs stellt die systematische und kontinuierliche Qualitätssicherung und -verbesserung der (nichtphilologischen) fremdsprachlichen Ausbildung für Studierende aller Fachrichtungen an der TU Dortmund dar.

Zu einer qualifizierten Fremdsprachenausbildung gehört nicht nur das Unterrichten, sondern auch das Prüfen, Bewerten und Zertifizieren. Ohne Leistungsmessung ist keine fundierte Aussage über fremdsprachliche Kompetenzen möglich. Universitäten müssen dafür Sorge tragen, dass ihre Absolvent*innen für den globalen Wissenschaftsbetrieb und Arbeitsmarkt fremdsprachlich optimal qualifiziert sind und dementsprechend die Vergabe von fundierten und aussagekräftigen Nachweisen vorsehen.

Auf welche Art und Weise können universitäre Sprachlehreinrichtungen fremdsprachliche Kompetenzen (*learning outcomes*) am besten objektivieren, sprich messen, beurteilen und bewerten? Der *Gemeinsame europäische Referenzrahmen für Sprachen* des Europarats (2001) samt seinen Skalen- und Kriterienkatalogen sowie dem GeR nachfolgende Publikationen wie beispielsweise *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR)* (European Council, 2009) oder jüngst *Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Companion Volume* (European Council, 2018) bieten selbstredend wesentliche Werkzeuge für die Beschreibung und Objektivierung von fremdsprachlichen Kompetenzen. Eine kontinuierliche Herausforderung für universitäre Sprachlehreinrichtungen und somit für unseren Bereich Fremdsprachen besteht allerdings darin, die im GeR eher unspezifisch und allgemein formulierten Deskriptoren für den Hochschulkontext sowohl im Unterrichts- als auch im Prüfungsgeschehen adressatengerecht auszulegen und anzuwenden.

Universitäre Sprachlehreinrichtungen werden in regelmäßigen Abständen mit der Frage konfrontiert, wie sich die eigenen internen Prüfungen von den standardisierten Englischtests internationaler Testanbieter wie TOEFL, Telc oder IELTS abheben bzw. welche Prüfungsergebnisse aussagekräftiger und (implizit) „qualitativ hochwertiger" sind. Schließlich sind diese standardisierten Tests international anerkannt und beruhen auf umfangreichen Kalibrierungen und Vorerprobungen – ein Umstand, den die meisten hochschulinternen Sprachprüfungen bzw. Sprachtests nicht erfüllen können. [1]

In diesem Zusammenhang ist das Hervorheben der Hochschul- und Berufsspezifik sowohl von Unterricht als auch von Prüfungen an universitären Sprachlehreinrichtungen wesentlich. Hochschulinterne (*ausbildungsbezogene*) Prüfungen bereiten Studierende sehr spezifisch auf das fremdsprachliche Agieren und Interagieren in hochschulischen und beruflichen Kontexten vor, während internationale (*ausbildungsunabhängige*) Prüfungen inhaltlich und fachlich ein eher allgemeines Themenspektrum aufweisen, dafür aber in hohem Maße hinsichtlich des Testformats standardisiert sind und nicht nur standortbezogen, sondern über nationale Grenzen hinweg die Sprachkompetenzen von Testteilnehmenden prüfen und zertifizieren. Die Zwecke, Zielsetzungen, Aufgabenformate und Qualitätssicherungsmaßnahmen von hochschulinternen Sprachprüfungen und standardisierten Tests kommerzieller Testanbieter unterscheiden sich voneinander. Unser Ziel ist es via Unterricht und Prüfung Studierende fremdsprachlich handlungsfähig zu machen für studien- und berufsbezogene Kontexte und deren Handlungsfähigkeit stufenadäquat valide zu beurteilen und zu bewerten. Das Ziel von internationalen Testanbietern besteht darin, unabhängig von einem bestimmten Ausbildungsprogramm eine valide und präzise Messung des Sprachstandes eines/r Testteilnehmenden vorzunehmen.

Obwohl ein hohes Maß an Standardisierung von hochschulinternen Sprachprüfungen auf der Basis von Kalibrierung und Vorerprobung an den allermeisten universitären Standorten aus Ressourcengründen nicht möglich ist, haben universitäre Sprachprüfungen ebenfalls testmethodischen Güte- und Qualitätskriterien zu entsprechen, müssen also objektiv, transparent, reliabel, valide und aussagekräftig sein. Wie gelingt dies erfolgreich? Eine Antwort des Arbeitskreises deutscher Sprachenzentren (AKS) auf diese Frage ist: Durch die Implementierung und kontinuierliche Optimierung handlungsorientierter Kurs- UND Prüfungskonzepte im universitären Fremdsprachenunterricht.

Was verstehen wir konkret unter „Handlungsorientierung"? Diese Begrifflichkeit ist im GeR fest verankert: Der Lernende wird eine als *in* und *mit* der Fremdsprache handelnde Person begriffen, die ein bestimmtes kommunikatives Ziel verfolgt. Der handlungsorientierte

[1] Die Begrifflichkeiten „Prüfung" und „Test" werden im fremdsprachendidaktischen Diskurs mitunter undifferenziert als Synonyme verwendet. Ich schließe mich im Verständnis der beiden Begrifflichkeiten Thomas Tinnefeld an, der in seiner Publikation *Dimensionen der Prüfungsdidaktik* (2013) *Prüfungen* als einen übergeordneten Begriff, sieht, der *Tests* miteinschließt: „Prüfungen sind *per definitionem* als holistisch zu betrachten, und zwar in dem Sinne, dass sie allumfassend sind und jede Situation betreffen, in der Wissen auf formelle und informelle Art und Weise […] zur Leistungsüberprüfung zwecks des Erwerbs von Qualifikationen oder auch zur Überprüfung des eigenen Lernstandes abgefragt wird. Standardisierte Tests […] sind lediglich ein potentieller Teil aller denkbaren bzw. real durchgeführten Prüfungsformen und somit nicht holistisch, sondern vielmehr partiell ausgerichtet. […] Es ergibt sich zwischen Prüfungen und Tests somit ein klares Inklusionsverhältnis. […] Allgemein kann festgestellt werden, dass nicht jede Prüfung ein (standardisierter) Test ist; dagegen ist jeder (standardisierte) Test eine Prüfung" (S. 119).

Fremdsprachenunterricht zielt darauf ab, fiktive Kommunikationssituationen durch *möglichst* realitätsnahe Kommunikationssituationen und Handlungskontexte zu ersetzen. Im Mittelpunkt des Unterrichts- und Prüfungsgeschehens steht die Bewältigung von authentischen Aufgaben und Problemen, die mit der Lebenswelt der Studierenden in enger Verbindung stehen (z.B. ein Exzerpt aus einem Fachartikel anfertigen, ein Sprechstundengespräch führen, eine Präsentation vor der Projektgruppe halten, einen Lebenslauf verfassen etc.). Die Lehrperson erfüllt hierbei weniger eine wissensvermittelnde Funktion, sondern übernimmt eher eine moderierende und beratende Rolle im Prozess der Aufgabenbewältigung. Kollaborative Arbeitsformen wie zum Beispiel Simulationen, Fallstudien oder Gruppenprojekte mit einer klaren differenzierenden Aufgabenteilung und einer möglichst studienbezogenen und/oder berufsrelevanten Themenstellung sind in diesen Unterrichtssettings besonders geeignet, lebensnahe Interaktionsprozesse unter den Lernenden in Gang zu bringen und zu authentischen Sprachhandlungen (die sich von den mehr oder weniger fiktiven oder semi-authentischen Sprachübungsaufgaben so mancher Lehrwerke abheben) zu motivieren.

Auf der Basis eines GeR-stufenadäquaten, handlungsorientierten Ansatzes, der authentische Themen und Fragestellungen aus Studienalltag und Berufskontext integral berücksichtigt, lässt sich in einer aussagekräftigen und validen Form überprüfen, „wie gut sich die Prüfungskandidatin oder Prüfungskandidat in einem bestimmten Kontext in einer konkreten Situation des Studienkontextes, des Fachstudiums oder des späteren Berufs in der Fremdsprache verständigen kann. Eine gut konzipiert handlungsorientierte Prüfung hat somit hohen Aktualitätsbezug, weist einen hohen Grad an Authentizität und Relevanz auf und ist somit valide“ (Fischer, 2013, S. 96).

Aus den oben dargelegten Rahmenbedingungen lässt sich für uns das folgende Fazit ziehen: Als Lehrende in einer universitären Sprachlehreinrichtung sind wir gleichzeitig immer auch Prüfende und in beiden Rollen wird von uns eine hohe Professionalität abverlangt. Wir sind nicht nur gefordert, unseren Unterricht GeR-stufenadäquat zu konzipieren und durchzuführen, sondern wir müssen ebenso darauf achten, die Konzeption, Durchführung und Bewertung von Prüfungen in Einklang mit dem GeR zu bringen. Gleichzeitig haben wir dafür Sorge zu tragen, dass das universitäre Prüfungsgeschehen grundlegende testmethodische Gütekriterien berücksichtigt, ohne dass eine Vorerprobung und Kalibrierung in gleicher Weise wie bei den standardisierten Tests erfolgen kann.

Zur Gewährleistung dieser Gütekriterien braucht es einen kontinuierlichen Dialog über die hochschulspezifische Auslegung der GeR-Deskriptoren sowie die Erstellung und Anwendung transparenter Checklisten und Bewertungsraster, die uns als Lehrende und Prüfende in der Prüfungskonzeption, Durchführung und Bewertung unterstützen. Solche Checklisten und Bewertungsraster stellen wir in unserem Moodle-Arbeitsraum *Englisch Organisation* zur Verfügung, der für alle Lehrkräfte des Lehrgebiets Englisch im Bereich Fremdsprachen zugänglich ist. Die Checklisten und Bewertungsraster sollen dazu beitragen, dass innerhalb des gesamten Lehrgebiets Englisch eine transparente und vergleichbare Handhabung in der Konzeption von Prüfungen sowie in der Bewertung von Prüfungsleistungen gewährleistet wird. Gleichzeitig sind sie als dynamische Werkzeuge zu verstehen, die wir regelmäßig an die sich verändernden universitären Gegebenheiten anpassen müssen.

Die primäre Zielsetzung des vorliegenden *Assessment Guide* ist es, zu einer vertiefenden Reflektion und einem fortgeführten Dialog über den Themenkomplex Prüfen und Bewerten im Lehrgebiet Englisch – und darüber hinaus im gesamten Bereich Fremdsprachen – anzuregen. Der *Assessment Guide* soll als Leitfaden für die Prüfungs- und Bewertungstätigkeit verstanden werden. Er ist bewusst nicht einstimmig, sondern mehrstimmig gestaltet, d.h. unterschiedliche Ansätze und Ansichten zum Themenkomplex Prüfen und Bewerten werden einbezogen. *Prüfen und Bewerten* wird in einem breiten Sinne verstanden: Die Begrifflichkeiten umfassen eine Palette unterschiedlicher Prüfungs- und Bewertungsformen, also nicht allein schriftliche Abschlussklausuren oder mündliche Abschlussprüfungen, sondern ebenso kumulative Formen der Leistungserbringung und -bewertung wie beispielsweise Portfolioarbeit. Sicherlich hat uns die durch die Corona-Pandemie bedingte „digitale Wende“ dazu geführt, in einem noch stärkeren Maße über unterschiedliche Prüfungs- und Bewertungsformen nachzudenken und diese in unseren Kursen zu implementieren.

Prüfen sollte in unserer Lehrpraxis keinen vorrangigen Stellenwert im Sinne eines *teaching-to-the-test* einnehmen. Dennoch sind Prüfungen im Fremdsprachenunterricht unentbehrlich, da sie eine verlässliche Aussage über Spracherwerb und Sprachkompetenzen erst möglich machen. Mithilfe eines der Hochschul- und Berufsspezifik Rechnung tragenden, handlungsorientiert gestalteten Unterrichts- und Prüfungsgeschehens können wir Studierende englischsprachig/fremdsprachig optimal auf spätere Berufsfelder am globalen Arbeitsmarkt vorbereiten.

Section 1: Assessment Basics nimmt den Begriff *assessment* näher in den Blick und stellt in diesem Zuge unterschiedliche Prüfungs- und Bewertungsformen dar. Des Weiteren fokussiert das Kapitel Reliabilität und Validität als zwei wesentliche testmethodische Gütekriterien und beschreibt Wege der Umsetzung dieser beiden Gütekriterien bei der Konzeption unserer Englischprüfungen.

Section 2: Designing End-of-Term Valid English Tests gibt praktische Hinweise und Tipps für die Konzeption von stufenadäquaten, validen Prüfungen in den vier Fertigkeitsbereichen auf den Niveaustufen B2 – C2. Thematisiert werden die Auswahlkriterien für Prüfungstexte sowie für Aufgabenformate hinsichtlich der Erstellung valider schriftlicher und mündlicher Semesterabschluss-Prüfungen.

Dem Portfolio-Ansatz widmet sich *Section 3: Portfolio-Based Assessment: Principles for Practice in an EFL Classroom*. Hier wird argumentiert, dass Portfolioarbeit eine alternative Prüfungs- und Bewertungsform darstellt, welche die Lernautonomie maßgeblich fördert. Portfolioarbeit bezieht Studierende stärker in Entscheidungsprozesse hinsichtlich der eigenen Lerngestaltung ein und nimmt sie dementsprechend stärker in die Pflicht als traditionelle Abschlussklausuren. Während Abschlussklausuren punktuelle Momentaufnahmen einer fremdsprachlichen Leistung darstellen, bieten Portfolios die Möglichkeit der Leistungsevaluierung über einen längeren Zeitraum.

Die DAAD-Prüfung steht im Fokus der *Section 4: Using Benchmarks to Assess Student English Language Proficiency Levels in the DAAD Exams*. Im Zuge der Bewerbung für ein Auslandsstudium oder ein Auslandspraktikum innerhalb der EU wird häufig der sogenannte

DAAD-Sprachnachweis verlangt. Diesen Nachweis können Studierende der TU Dortmund im zhb Bereich Fremdsprachen erwerben. Der Abschnitt beschreibt das Format unserer DAAD-Prüfung und demonstriert überdies – in exemplarischer Weise für den mündlichen Ausdruck – die Konzeption von niveauspezifischen *benchmark*-Beschreibungen, die für eine valide Bewertung der Prüfungsleistungen zweckdienlich sind.

Section 5: Possible Summative Assessment Tasks and Activities at the B1 to C2 Levels und *Section 6: Tips and Strategies for Self-Editing Your Tests* bieten eine Übersicht über geeignete Aufgabentypen und Aufgabenstellungen.

An dieser Stelle möchte ich Tetyana Müller-Lyaskovets, Kai Herklotz und Geoff Tranter für die aussagekräftigen Beiträge zu diesem *Assessment Guide* sehr danken. Ein herzliches Dankeschön geht zudem an Helen Horner von der *University of Minnesota* (USA) für das detaillierte Lektorat dieser Publikation. Mein ganz besonderer und ausdrücklicher Dank richtet sich an alle Lehrbeauftragten in unserem Lehrgebiet Englisch, die den Hauptteil des Lehr- und Prüfungsgeschehens schultern und sich kontinuierlich mit großem Engagement dafür einsetzen, dass Studierende der TU Dortmund eine hochwertige akademische Englischausbildung erhalten.

Referenzen

Publikationen European Council:

Common European Framework of Reference for Languages. Learning, Teaching, Assessment (2001). Strasbourg: European Council. https://rm.coe.int/1680459f97

Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Companion Volume (2020). The European Council.
https://www.coe.int/en/web/education/-/common-european-framework-of-reference-for-languages-learning-teaching-assessment-companion-volume

Relating Language Examinations to the 'Common European Framework of Reference for Languages: Learning, Teaching, Assessment' (CEFR). A Manual (2009). The European Council.
https://www.coe.int/en/web/common-european-framework-reference-languages/relating-examinations-to-the-cefr

Fremdsprachendidaktische Publikationen:

Fischer, J. (2013). Zum Verhältnis von standardisierten und ausbildungsbezogenen Sprachtests im Hochschulkontext. In: Gekonnt, verkannt, anerkannt? – Sprachen im Bologna Prozess. Hrsg. Regina Mügge. Dokumentation 13. AKS Verlag.

Tinnefeld, T. (2013). *Dimensionen der Prüfungsdidaktik. Analysen und Reflexion zur Leistungs-bewertung in den modernen Fremdsprachen*. Saarbrückener Schriften zu Linguistik und Fremdsprachendidaktik. Bd.1.

Vogel, T. (2009). Eröffnung der 25. Arbeitstagung des AKS an der Universität Passau 2008. *Sprachen als akademische Schlüsselkompetenz?* Dokumentation 11. AKS Verlag.

Section 1: Assessment Basics

Tetyana Müller-Lyaskovets

Why do we assess our students? Why do we favor tests in assessing our students' language learning outcomes? How good are our language tests and other assessments? Do we assess proficiency levels, achievement, or both? English language instructors driven by the examination culture of our universities, we have become so accustomed to measuring our students' performance that we tend to forget asking important questions about the purposes and quality of our assessments. However, our assessments matter in so many significant ways. They have a direct impact on our teaching practices and our students' learning experiences and outcomes. They matter institutionally – in terms of how our departments participate in larger conversations about language center practices, accountability, and funding. Finally, our assessments speak to our professionalism and teaching philosophies.

The purpose of this Assessment Guide is to offer recommendations and guiding templates to be used by the EFL (English as a Foreign Language) instructors in the design and development of valid, reliable, and fair assessments that enhance students' learning experience and learning outcomes. The Assessment Guide aims to promote principles of consistency of practice across all sections of CEFR-informed English instruction in the Department of Foreign Languages, TU Dortmund University.

What is Assessment?

Definition

Over the years, education sciences have generated multiple definitions of assessment and its role in teaching and learning. We offer two definitions that best capture how we envision assessment in our department.

Assessment "refers to the wide variety of methods or tools that educators use to evaluate, measure, and document the academic readiness, learning progress, skill acquisition, or educational needs of students" (The Glossary of Education Reform).

Assessment is "the systematic basis for making inferences about the learning and development of students. It is the process of defining, selecting, designing, collecting, analyzing, interpreting, and using information to increase students' learning and development" (Erwin, 1991, p. 14).

Based on current popular practices at German university language centers, instructors tend to think of assessment as carried out via **end-of-course** or **standardized language tests**. However, assessment is not reduced to these test formats. Some other formats and methods exist that measure student language performance and proficiency. Each of these other assessment methods has its own utility and context. Moreover, an assessment activity may fall into several categories. **Placement assessments** are administered to determine whether a student is ready to participate in a language course designed for a specific level. We use

summative assessment to evaluate our students' learning outcomes at the end of a certain period such as, for example, at the end of the semester. Unlike summative assessment that culminates in giving students a grade, **formative assessment** has a diagnostic and remedial function. The purpose of formative assessment is to diagnose a problem and to suggest ways to improve. According to Black and Wiliam (2009), "Practice in a classroom is formative to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited" (p. 9). Black and Wiliam (2009) suggest the following formative assessment strategies: sharing success criteria with learners, classroom-only questioning, comment-only marking, peer- and self-assessment, and formative use of summative tests (p. 7). Unlike traditional tests, **performance assessments** ask students to work on **an authentic task**, such as, for example, writing a paper, giving a talk, preparing a presentation, or completing a project (The Glossary of Education Reform). Finally, **portfolio-based assessments** require students to select their work completed over some time to demonstrate their skills and achievements.

Very often, assessments are described as **proficiency assessments** or as **achievement assessments**. For example, such standardized tests as the TOEFL or Cambridge tests are usually referred to as "proficiency" tests. In our context, to assess achievement means to assess what has been taught in the course, whereas to assess proficiency means to evaluate student language proficiency with or without regard to a particular course. Because our priority is to provide high quality language instruction and student learning experience, we are challenged to reconcile both approaches. We want to test achievement and performance at a specified proficiency level, such as, for example, B2, C1, or another level. Performing pure proficiency assessments would be beyond the scope of courses offered by German university language centers. High-quality proficiency assessments require more resources than we can afford, such as, for example, statistical analysis or piloting, to name just a few.

When choosing an assessment method, first of all, we need to think about what we want to achieve through this assessment.

What is the Purpose of Assessment?

In defining the **purposes of assessment**, we adopt a student-centered approach that takes into account not only institutional interests but also students' interests and gains. In thinking about assessments, instructors should also ask the question of how their students will benefit from these assessments. Students should be perceived as important stakeholders of assessments that are used to improve student learning experience rather than as a punitive force of instruction.

Clay and Root (2001) note that we assess our students to achieve the following:

- "to provide a record for assigning grades
- to provide a learning experience for students
- to motivate students to learn

- to communicate to students their level of understanding of the course objectives and… [provide students with] a guide for further study
- to assess how well students are achieving the stated goals and course objectives
- to provide the instructor with an opportunity to reinforce the stated objectives and highlight what is important for students to remember" (p. 52).
- to obtain data that can be used as a teacher guide for further course improvement.

What Can I Do to Increase Assessment Reliability?

Reliability is one of the basic constructs in measurement theory and a vital criterion for developing assessments. Reliable assessment is a *consistent and reproducible* assessment. Imagine that two instructors are grading independently the same essay. Will their assessments yield the same results?

The assessment results can be consistent, but up to a certain degree, because a true score exists only in theory. In real life, we obtain a student's *observed score*. This *observed score* consists of a *true score and an error*. The error can be caused by the following factors: student characteristics, test characteristics, and the conditions that influence the administration and scoring of the task (Cherry & Meyer, 2009, p. 31). It is important to remember that assessments are not laboratory experiments and that their quality should be viewed through the lens of several *reliability types*.

Interrater reliability measures agreement among the instructors performing the same assessment. We should take this measure into account if we want to minimize subjectivity. *Parallel forms reliability* measures equivalence. We should take this measure into account when, for example, creating a new assessment for the same course each new semester. If the same student was to take a reading comprehension (RC) test and then take a new version of an equivalent RC test designed for the next semester, will this student show similar results? *Internal consistency reliability* shows if each item and element of the assessment measure the same construct. One can use a multiple-choice item as an example: each option in the item must measure the same construct.

In their ETS publication, Young, Youngsoon So, and Ockey (2013) offer the following definition of reliability: "**Reliability** refers to the extent to which an assessment yields the same results on different occasions. Ideally, if an assessment is given to two groups of test-takers with equal ability under the same testing conditions, the results of the two assessments should be the same, or very similar" (p. 5).

What can we do to make more consistent judgments about our students' abilities? Researchers and practitioners usually agree on the following measures:

- Use enough items, questions, or tasks to assess a competence – the more the better. Please see Pope's (2010) posting for a more detailed explanation.

- Make sure that your students are familiar with the assessment format and grading criteria.

- Make sure that assessment environment is the same for each participant and that this environment does not underprivilege any of your students. For example, each student has the same amount of time to complete a test or to write a take-home essay.
- Create grading rubrics that help instructors apply the same criteria across all graded student work.
- Grade the first item/task of each student paper in a batch and then move to grading the second item/task of each student paper in the same batch. This process will not only help you increase the reliability of the results, but it will also save you a lot of time spent grading.
- Check your tests consistently for ambiguous and poor-performing items/tasks. Revise your assessments.
- Grade anonymously to avoid teacher biases. Imagine how you would react to a test paper of a hard-working or motivated student versus a test paper of a less motivated or less diligent student.

As a psychometric construct, reliability is viewed as a necessary but insufficient condition for validity (Cherry & Meyer, 2009, p. 30). Assessments can be consistent but not valid, which also means that we can do a consistently poor job of assessing our students' abilities and achievement. Our assessments need to satisfy more criteria to be valid.

What Can I Do to Increase the Likelihood of Developing a Valid Assessment?

Validity comes from the word "validate," that is to provide evidence that supports both interpretations and uses of assessment results. The classical positivist definition of validity asks the following question: "Does the test measure what it is supposed to measure? If it does, it is valid" (Lado, 1961, p. 321). This approach to validity reflects psychometric testing practices that privilege accuracy and truth over context and the measure's value. However, since the end of the 20th century, we have been witnessing a shift in focus from validating the test *to validating the test's uses and its impact on learning*: "Validity as it is currently understood is about validating decisions based on an assessment" (An Introduction to Writing Assessment Theory and Practice, 2009, p. 4). This latter approach to assessment urges us to ask questions about a relationship between our teaching practices and assessments, about what is it that we are teaching and assessing, and about how we want to use our assessments and their results. It urges us to contextualize the assessment in a current local situation and consider ethical implications of both result interpretations and their uses. Thus, the current understanding of validity incorporates not only *construct validity* (What is it that we are testing?), but it also includes *value implications, assessment's relevance and utility, and assessment's social consequences* (Chapelle, 1999, p. 259).

Although there are different types of validity, literature on validity portrays *construct validity* as central to our understanding of validity. The following table summarizes the transformation that understanding of validity has undergone:

Table 1. Summary of Contrasts Between Past and Current Conceptions of Validation (Chapelle, 1999, p. 258)

Past	**Current**
Validity was considered a *characteristic of a test:* the extent to which a test measures what it is supposed to measure.	Validity is considered an *argument* concerning test interpretation and use: the extent to which test interpretations and uses can be justified.
Reliability was seen as distinct from and a necessary *condition for validity.*	Reliability can be seen as *one type of validity evidence.*
Validity was often established through *correlations* of a test with other tests.	Validity is argued on the basis of a number of types of *rationales and evidence,* including the consequences of testing.
Construct validity was seen as one of *three types of validity* (the three validities were content, criterion-related, and construct).	Validity is a *unitary concept* with construct validity as central (content and criterion-related evidence can be used as evidence about construct validity).
Establishing validity was considered within the purview of *testing researchers* responsible for developing large-scale, high-stakes tests.	Justifying the validity of test use is the responsibility of *all test users.*

Validation is a complex process that requires resources and validation expertise. Most of the language centers in Germany do not have these resources. What is it that we as teachers can do to increase the likelihood of developing valid assessments?

Practical tips for the Departments of Foreign Languages to increase validity of their assessments:

To make valid interpretations based on the assessment results, we can look at the content of our assessments in relation to assessment specifications. To do so, first, we need to **create test specifications** that will clearly define

- the purposes of assessment, including the constructs and content that it targets
- the target population.

Test specifications for the courses offered by the Department of Foreign Languages, TU Dortmund University are available for download from the Moodle Platform for English instructors.

To develop valid assessments, we have to consider the ways our students respond to and perform on our assessments. In addition, we can collaborate with other instructors and institutions to create better assessments. Finally, we must take into account not only positive but also negative side effects of our assessments, negative washback among others, such as teaching to the test. Thus, for example, test preparation, as researchers argue, may have "a negative influence on teachers' instruction due to a focus on procedural skills" (Blazar & Pollard, 2017, p. 420). In addition, test preparation activities were found to predict a lower quality of instruction for some classrooms (Blazar & Pollard, 2017, p. 420).

When thinking about validity of our assessments, it is important to define our positions and worldviews. In other words, do we prefer holistic or analytic psychometric scoring? In the context of assessing writing, for example, we need to reflect on whether we can objectively measure student skills and achievement through tests or rather we would like to create assessments that focus more on “the writer, writing process, and the development of the higher-level composing skills, such as self-reflection” (Silva, 1990, as cited in Lam, 2016).

Another important decision would be about values that we attach to a particular type of assessment. Would we like to test integrated skills or we would rather favor tests of discrete knowledge and skills, such as, for example, listening or speaking. Do presentations as an assessment activity provide a snapshot of speaking skills only or do they assess communicative competencies, given the fact that nowadays most of the presentations have a visual rhetoric and writing component?

Whatever assessment approach we choose, we need to remember that the perfect score does not exist and that tests and assessments yield only approximations rather than a “true” picture of our students’ ability levels. One way to address this predicament would be to combine several assessment activities in one language course.

One possible and perhaps the oldest way to check the quality of our test would be to evaluate its criterion validity. Imagine that we want to test our students’ reading skills at a certain CEFR level. First, we need to find an existing reading comprehension test that measures the same construct at the same level and, ideally, for the same purpose and whose validity has been recognized (for example, suitable TOEFL or Cambridge English tests) and have our students take both tests – the other test and our test. Then, we need to compare how our students performed on both tests, and if the results are very similar, we can conclude that our test has high criterion validity. Yet, we should remember that validity is not reduced to criterion validity, and it is a context-specific unitary concept.

Conclusion

Thinking about our assessments should lead us to the development of ethical standards that we will apply to our teaching, assessments, and educational environments. We invite you to be mindful of how you develop your assessments and what you strive to achieve through them. Practice your assessments thoughtfully, collaborate with other instructors in developing your assessments, and reflect on how your assessments affect your students and you. Think about the ethical and social implications of your assessments, of their intended and unintended consequences, such as the negative washback effect or the extent to which testing causes undue anxiety. Does your assessment improve your students’ learning experiences and create a positive washback into the curriculum or it rather takes control of your teaching by making you turn your teaching into a prepping exercise?

We hope that this assessment guide will help you make choices about assessment activities in your language classrooms and adjust your assessments to your teaching objectives, teaching contexts, and students’ needs. This guide includes contributions that not only highlight the problems and challenges of creating reliable and valid assessments but also provide practical recommendations for improving test development, implementing portfolio assessment, and using benchmarks to assess student English language proficiency levels in the DAAD exams.

References

An introduction to writing assessment theory and practice (2009). In B. A. Huot & P. O'Neill (Eds.). *Assessing writing: A critical sourcebook* (pp. 1-9). Bedford/St. Martins.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5-31. https://doi.org/10.1007/s11092-008-9068-5

Blazar, D., & Pollard, C. (2017). Does test preparation mean low-quality instruction? *Educational Researcher, 46*(8), 420-433. https://doi.org/10.3102/0013189x17732753

Chapelle, C. A. (1999). Validity in language assessment. *Annual Review of Applied Linguistics*, *19*, 254–272. https://doi.org/10.1017/s0267190599190135

Cherry, R. D., & Meyer, P. R. (2009). Reliability issues in holistic assessment. In B. A. Huot & P. O'Neill (Eds.). *Assessing writing: A critical sourcebook* (pp.29-56). Bedford/St. Martins.

Clay, B., & Root, E. (2001). *Is this a trick question? A short guide to writing effective test questions*. Kansas Curriculum Center.

Erwin, T. D. (1991). *Assessing student learning and development: A guide to the principles, goals, and methods of determining college outcomes*. Jossey-Bass.

Pope, G. (2010, May 10). *How many questions do I need on my assessment? Questionmark*. https://www.questionmark.com/how-many-questions-do-i-need-on-my-assessment

Lado, R. (1961). *Language testing: The construction and use of foreign language tests.* McGraw-Hill.

Lam, R. (2016). Assessment as learning: Examining a cycle of teaching, learning, and assessment of writing in the portfolio-based classroom. *Studies in Higher Education*, *41*(11), 1900–1917.
doi: 10.1080/03075079.2014.999317

The Glossary of Education Reform (n.d.). https://www.edglossary.org

Young, J. W., Youngsoon So, & Ockey G. J. (2013). *Guidelines for best test development practices to ensure validity and fairness for international English language proficiency assessments*. Educational Testing Service.
https://www.ets.org/s/about/pdf/best_practices_ensure_validity_fairness_english_language_assessments.pdf

Section 2: Designing Valid End-of-Term English Tests

Geoff Tranter

Foreword

The area of language-testing theory and the practical aspects of how to design reliable, valid, and relevant language tests covers such a wide range of important background information that it is impossible to incorporate everything into one chapter of these guidelines. For this reason, this chapter will concentrate more on those features that are of immediate relevance for examinations within the remit of the English courses offered by the Department of Foreign Languages at TU Dortmund University, i.e. English for Specific Purposes (ESP) and English for Academic Purposes (EAP) courses at CEFR levels of B1 and higher. Teachers interested in a more comprehensive insight into questions concerning general aspects of language testing are recommended to consult the following websites that offer more in-depth background information:

https://www.britishcouncil.org/exam/aptis/research/projects/assessment-literacy/general-language-proficiency
https://www.teachingenglish.org.uk/article/proficiency-test
https://www.cambridgeenglish.org/research-and-validation/about-language-testing

Introduction

The very first question language teachers need to ask themselves when confronted with the task of devising a test that will provide their students with a formal qualification at the end of the course is actually quite basic in nature, namely "What is the purpose of the test?" or "What does the test aim to give evidence of?" The question itself may seem relatively simple and obvious, but when answering this question, we first have to consider the needs of all the **stakeholders** involved:

- First of all, **the students**, whose main interest is not only to have feedback concerning their language proficiency, but also to have concrete evidence of that proficiency to include in job applications as part of the curriculum vitae;
- Secondly, **the potential employers**, who wish to have a well-founded and clearly understandable assessment of the student's language skills in order to be able to predict the applicant's ability to fulfil those tasks that the successful applicant will have to fulfil;
- Thirdly, **ourself as the teacher**, as we are interested in an evaluation of both our teaching outcomes and the students' learning throughout the course;
- Finally, **the language centre**, which needs to have feedback regarding the learning outcomes in the courses that the centre offers, and thus receive an overview of the quality of the language programmes it provides.

From this basic definition of the purpose of end-of-course testing can be deduced a number of important and relevant principles or quality standards that have to be considered in order to meet the needs of all the afore-mentioned stakeholders.

Types of Tests

Achievement Tests versus Proficiency Tests

As its name implies, the primary aim of an Achievement Test is to evaluate the degree to which the students have achieved the aims of a course in terms of the skills and content on which the teaching has been based, i.e. on the curriculum. A Proficiency Test, on the other hand, aims to ascertain whether the students have reached the level of language proficiency that the course was targeted at. As a result, end-of-term proficiency tests should be based on the skill definitions provided by the Common European Framework of Reference (CEFR) for the level the course was set at, and NOT be simply restricted to the content of the course.

→ *The end-of-course test should primarily be a proficiency test covering the skills as defined in the CEFR for the level of the individual course, not just within the framework of the content of the course.*

Tests of Skills versus Tests of Knowledge

As the main aim of language teaching is to train the students to achieve successful and effective communication, the primary content of any end-of-course test should focus on **the ability to use language** (= language skills) and **NOT to test the student's knowledge** of elements of the language (e.g. individual lexical or grammatical items).

→ *The end-of-term test should primarily focus on evaluation of the five skills (Speaking, Writing, Reading, Listening, and Mediation) as defined in the CEFR for the level the course was designed to achieve, and not on assessing knowledge of individual language elements*.

General Principles regarding the Quality of Tests

Context

As language for the purpose of communication is normally used within a context, any test that is intended to give evidence of the student's ability to use language in order to communicate effectively should be embedded in an appropriate context. This principle applies to all forms of language communication: both the productive skills of speaking and writing and the receptive skills of reading and writing.

Context consists of the following components:

- Roles: who is communicating with whom?
- Purpose: What is the purpose of the communication? What is the intended outcome?
- Setting/Situation: Where is the communication taking place?

The more detailed the context, the easier it becomes for candidates to understand and identify with the task. This prevents them from having to spend more time on working out what is actually expected of them and thus enables them to concentrate more on the language they need to deal with the task.

→ *All tasks in the end-of-course test should provide the candidates with full clarity regarding the background to the task, i.e. Who? Why? What? Where? How?*

Authenticity

If a test is intended to provide evidence of the students' ability to deal with tasks and assignments that they will be confronted with outside the classroom or the examination room, the test should always be based on test items that as far as possible reflect authentic, real-life situations.

Consequently, in testing, the principle of authenticity applies to

- the **context** in which the task is embedded
- the **role** of the student in the task
- the **setting** for the task
- the **texts** (written, spoken, video) used in the tasks should not be teacher-generated texts
- the **task** itself should be a task that the student can identify with.

→ *All tasks designed for language testing should have a basis in real-life situations, enabling the candidates to immediately identify with the task and recognise the relevance for real-life language communication.*

Designing Language Test Materials: Listening Comprehension

Texts and Tasks

When preparing a Listening Comprehension Test, the first basic decision that has to be taken is the question as to which of the three main listening strategies – listening for details or intensive listening, listening for gist or global listening, and "scanning" or listening for just one or two specific details – the test questions are going to focus on. **As a general rule, at least two of these three strategies should be included in order to cover a range of listening skills**. This decision will not only influence the choice of listening task, but it will equally and perhaps more importantly form the basis for the choice of text, as there is normally a direct connection between the type of text and the listening strategy we normally apply. For example, hardly anybody would listen to a weather report to find out details of the weather throughout the whole of Europe (intensive listening for details) or just for a general impression of what the weather is going to be like (listening for gist). We normally listen to weather reports in order to find out what the weather is going to be like in the local area or the area where we will be travelling to today or tomorrow (listening for a specific detail).

→ *In other words, first decide the listening strategy to be tested and then choose the appropriate type of text.*

Further questions to be considered are:

- Should the text be a monologue (one speaker) or a dialogue (a minimum of two speakers)?
- Should the text be in different accents, e.g., British or American, first-language speaker or second-language speaker?
- Should the texts be short or long or of different lengths (particularly important in view of the different listening strategies)?
- Texts should ideally have been produced for a general audience (e.g., not for language teaching purposes such as language course books).

→ *For a more comprehensive assessment of overall listening proficiency, the test should include at least one dialogue and one monologue, two different accents, and one short and one long text. Listening texts should be authentic in that they were not produced for teaching or testing purposes, but for real-life communication.*

Types and Sources of Texts

For authentic listening situations at B1 level and above, the most appropriate texts are podcasts, radio reports and interviews; and a wide variety is available (cf. below).

→ *It is often tempting to use videos that are available on YouTube and many other sites. Audio-Visual Listening Comprehension is an important skill that is one of the sub-skills included in the CEFR. However, in many cases, such videos are unsuitable for listening comprehension purposes because the script often relies on the pictures that are shown, and without access to the pictures, it becomes more difficult for the candidates to follow the text*.

For this reason, it is recommended to only use videos when the test venue is technically equipped for showing videos and the connections are reliable.

There are a number of websites offering free recorded material. Many recordings are unfortunately too long for testing (and teaching) purposes. However, interesting material can be downloaded and, if necessary, edited to the desired length by downloading the freely available software *Audacity (https://audacity.en.softonic.com/download.com)*, which is quite easy to use. If help is needed, there are a number of good training videos on YouTube.

The following sites can be recommended for use in the classroom:
- www.scientificamerican.com (US)
- www.nakedscientists.com (UK)

These sites offer a range of podcasts of varying lengths, from texts of two minutes' duration up to reports and interviews of about thirty minutes' length. Despite their names, these websites both offer not only scientific and technical material, but also general input on a whole range of topics suitable for a non-technical audience. Just enter a key word into the search engine of the website. A further advantage is that many of the recordings offer tapescripts.

The following site offers reports of between three to seven minutes' length on virtually all topics of general interest: https://www.bbc.co.uk/programmes/p025ht2f/clips

Other sites offering shorter texts can be found by googling "sixty second podcasts". For specific areas, google "Short Business Podcasts" or "Short Technical Podcasts" or "Short …. Podcasts." You will be rewarded with a wide variety of useful and interesting material.

→ *When using texts from any of these sources, it is essential to include the source.*

Length and Number of Texts

In most internationally recognised tests from the CEFR Level B1 upwards, a Listening Comprehension Test normally takes about 25-30 minutes, which includes instructions, pauses for reading the questions, and a second hearing of each text.

This means that the recordings to be provided should have a total length of around 10-12 minutes. These 10-12 minutes should be used to provide at least two if not three texts in order to have at least one short text and one long text. In B1 groups, it is advised not to have texts longer than five minutes in view of the lower listening proficiency at that level. In higher groups, the texts can go up to 7-8 minutes, e.g., for B2, two two-minute podcasts and one seven-minute recording might be used.

Choice of Topic

The choice of topic will depend on the overall aim of the course. In the case of ESP courses, e.g., Engineering, Business Communication, etc., the topic needs to be within the thematic area of the course. However, there is a risk that specialised knowledge will interfere with the aim of the test insofar as the students might be able to answer questions more or less on the basis of the knowledge without reference to the text, and certain students might then be discriminated against due to the fact that their specialised knowledge is in a different area. The following ideas have proved themselves in practice:

Choose a topic that is:
- very up-to-date concerning the very latest developments or looks into the future
- very general in nature, e.g., a critical analysis, a review
- very specific in terms of one particular project, experiment, company, etc.

Instructions for Listening Comprehension Tests

Even though a test in the normal sense of the word is not "authentic listening" as the candidates are only listening in order to pass the test, it is important to attempt to achieve as high a degree of perceived authenticity (cf. above) as possible. This perceived authenticity can be achieved by providing a detailed context (cf. below) in order to enable the students to understand the listening situation as clearly as possible. It needs to be remembered that students are unable to ask any questions during the test – a difficult situation in which any uncertainty can distract from the actual task in hand.

Here is an example of a clear and useful instruction containing all the relevant information that students need in order to be able to focus on the text and questions.

"You① are now going to hear a two-minute② podcast③ (www.scientificamerican.com) about bridge design④. You are interested in this topic and would like more details⑤. Listen to the recording and note down the answers (complete sentences are not necessary⑥) for Questions 4 – 7⑦.
You will hear the text twice⑧. You now have 45 seconds⑨ to read Questions 4 – 7.
① Your role
② Length of recording
③ Type of text
④ Topic
⑤ Reason for listening
⑥ How much to write
⑦ Which questions to answer
⑧ Number of times
⑨ Preparation time

Listening Comprehension Tasks

There are basically three standard test formats that can be used to test Listening Comprehension: Multiple-Choice, True/False, and Open Questions.

Multiple-Choice Tasks
Multiple-Choice tasks are frequently used in testing situations where there are large numbers of candidates, as these tests need far less marking time and thus offer a high degree of economy, especially when they can be computer-marked. Despite that obvious advantage, multiple-choice tasks involve a number of inherent issues.

- Firstly, they are not easy to develop. The items must not be obviously right or wrong, and individual items should not cancel each other out. It must not be possible to recognise the correct answer based on worldly knowledge without reference to the listening text, and the items should be as short as possible.
- Secondly, by their very nature, multiple-choice items involve a certain degree of Reading Comprehension, all the more so if the items are long, and, if not completely clear, they can thus distort the validity of a Listening Proficiency test.
- Thirdly, multiple-choice tasks are not in any way authentic in term of listening skills. Outside a classroom or examination room, nobody listens to a text with three (or sometimes) four possible outcomes in mind.

True/False
The True/False format is another way of assessing listening proficiency and is frequently used in international examination systems as it is also easy to mark quickly (test economy). There is still the disadvantage that this format involves a certain degree of reading comprehension, but less so than with the multiple-choice format, as there is only one statement to read.

The main possible disadvantage of True/False listening comprehension items is the fact that there is usually a 50:50 chance of guessing the correct answer. For this reason, a more valid method of assessing listening skills is the **'True/False with evidence from the text'** format, where candidates have to quote from the text (no more than several words) to justify their answers.

Open Questions/Note-Taking
Although this format does not fulfil the criterion of test economy and takes more time to mark, it goes much closer to fulfilling the criterion of authenticity, which from the perspective of test validity is probably more important. Just as in real life when the candidates listen to the text to find out the information they want/need, in the case of the test, they listen to the text to identify the details they need to answer the questions.

There are two types of Open Questions, both of which involve taking notes on what is heard in the recording:

Actual open questions, e.g., What information is given about (bridge design in the US)? (one point)

#____

Sentence completion, e.g., Commenting on bridge design in the US, the speaker says (one point)

#____

Designing Language Test Materials: Reading Comprehension

Similar to Listening Tests, the same basic decision has to be taken when creating Reading Comprehension Tests, namely the question of which of the three main reading strategies – reading for details or intensive reading, reading for gist or global understanding, and "scanning", i.e. reading for just one or two specific details – the test questions are going to focus on. **As a general rule, at least two of these three strategies should be included in order to cover a range of reading skills**. This decision will not only influence the choice of reading task, but it will equally and perhaps more importantly form the basis for the choice of text, as there is normally a direct connection between the type of text and the reading strategy we normally apply. For example, it is unlikely that people would read an instruction manual in detail from the first to the last page (reading for detail or intensive reading). They are more likely to go through the manual as quickly as possible until they find the section and information they need ("scanning").

→ *In other words, first decide the reading strategy to be tested and then choose the appropriate type of text.*
Further questions to be considered are:

- Should the texts be short or long or of different lengths (particularly important in view of the different reading strategies)? As from the B2 level, where the CEFR reading proficiency descriptors refer to “long” and “lengthy” texts, texts of two pages or more can be used.

- The question of the appropriate length of a text is also connected with the density of a text, which refers to the density of content (the number of ideas expressed) and the lexical density (the number of advanced vocabulary and complex sentences). Depending on how long the candidates are given to complete the tasks, both the length of the text (in terms of number of words) and the density of the text need to be considered.
- The texts should have been produced for a general audience (e.g., not purely for language teaching purposes such as language course books).

Choice of Topic

The choice of topic will depend on the overall aim of the course. In the case of ESP courses, e.g., Engineering, Business Communication, etc., the topic needs to be within the thematic area of the course. However, there is a risk that specialised knowledge will interfere with the aim of the test insofar as the students might be able to answer questions more or less on the basis of the knowledge without reference to the text and certain students might then be discriminated against due to the fact that their specialised knowledge is in a different area.

The following ideas have proved themselves in practice. Choose a topic that is:
- very up-to-date concerning the very latest developments or looks into the future
- very general in nature, e.g., a critical analysis, a review
- very specific in terms of one particular project, experiment, company, country, etc.

Types and Sources of Texts

For authentic reading situations at the B1 level and above, the most appropriate texts are reports and newspaper articles, also blogs, transcripts of radio reports and interviews; and a wide variety is available (cf. below).

The following sites can be recommended:
- https.//www.scientificamerican.com (US)
- https://www.nakedscientists.com (UK)

Both sites offer a range of reports and articles of varying lengths. Despite their names, these websites offer both scientific, technical material and general articles on a whole range of everyday topics suitable for a non-technical readership.

For newspaper articles, https://www.theguardian.com/uk offers free access.
Another very useful site is https://www.bbc.com. Especially the sections *Future*, *Worklife*, and *Travel*, which offer reports of varying length on virtually all topics, might be of interest.

For specific areas, google “Business Blogs” or “Technical Blogs” or “(topic) Blogs.” You will be rewarded with a wide variety of useful and interesting material.

→ *When using texts from any of these sources, it is essential to include a reference to the source.*

In most internationally recognised tests as from the CEFR Level B1 upwards, a Reading Comprehension Test takes about 60 minutes. Experience shows that at the B1 level, about 2 – 2½ pages of text are an appropriate length, at the B2 level, 3 – 4 pages, and at the C1 level, 4 – 5 pages are an appropriate length.

Designing Reading Comprehension Tasks

There are basically three standard formats that can be used to test Reading Comprehension: Open Questions, True/False, and Multiple-Choice.

Multiple-Choice Tasks
As already mentioned in connection with Listening Comprehension, Multiple-Choice tasks are frequently used in testing situations where there are large numbers of candidates, as these tasks need far less marking time and thus offer a high degree of economy, especially when they are computer-marked. Despite that obvious advantage, they involve a number of inherent issues.

- Firstly, they are not easy to develop. The items must not be obviously right or wrong and individual items should not cancel each other out. It must not be possible to recognise the correct answer based on worldly knowledge without reference to the reading text, and they should be as short as possible.
- Secondly, they increase the reading load for candidates. Not only do the candidates have to understand the text, but they also have to focus their attention on a precise understanding of the possible answers, especially in the case of four-part multiple-choice items. Failure to understand the distractors, especially if the items are long or not completely clear, can distort the validity of the test, as failure in the reading part of the test may be a result of the quality of the items rather than a candidate's lack of comprehension of the text.
- Thirdly, multiple-choice items are not in any way authentic in term of reading skills. Outside the classroom/examination room, nobody reads a text with three (or sometimes) four possible outcomes in mind.

Open Questions/Note-Taking
Although this format does not fulfil the criterion of test economy and takes more time to mark, it does go closer to fulfilling the criterion of authenticity, which from the perspective of test validity is probably more important. Just as in real life when the candidates read the text to find out the information they want/need, in the case of the test, they read the text to identify the details they need to answer the questions.
True/False
The True/False format is another way of assessing reading proficiency and is frequently used in international examination systems as it is also easy to mark quickly.

The main possible disadvantage of True/False reading comprehension items is the fact that there is usually a 50:50 chance of guessing the correct answer. For this reason, a more valid method of assessing reading skills is the **'True/False with evidence from the text'** format, where candidates have to quote from the text (no more than several words) to justify their answers (cf. example below).

Instructions for Reading Comprehension Tests

Even though a test in the normal sense of the word is not "authentic reading," it is important to attempt to achieve as high a degree of perceived authenticity as possible. This can be achieved by providing a clear context in order to enable the students to understand the reading situation as far as possible.

Here is an example of a clear and useful instruction that contains all the relevant information that students need in order to be able to focus on the text and questions:

*"You are interested in the topic of storing energy*➀ *and find the following newspaper article* ➁ *(https://www.theguardian.com/environment)*➂**.** *Read the article and then do tasks 1 – 12*➃ *which you will find at the end of the text*➄*. You have sixty minutes*➅ *to read the text and find the answers."*

① The topic
② The type of text
③ The source of the text
④ The task
⑤ Where the tasks can be found
⑥ Time available
⑦ Precise task
⑧ Marks awarded
⑨ Example to make task clear

*"Now look at questions 1 – 12. Decide from your reading of the text if the statements are true or not true*➆ *according to the information in the text. Then give evidence from the text to support your answer. You will receive two points*➇ *for each correct answer* ***with evidence from the text****.*

Example➈: We do not have reservoirs for electricity.
[X] True, because we don't have the same system for electricity.
[ ] Not true, because ____
01. There are many alternatives to the traditional way of storing energy.
    [ ] True, because ____
    [ ] Not true, because ____
02. One leading firm is optimistic about the immediate future for storing energy.
    [ ] True, because ____
[ ] Not true, because ____

Designing Language Test Materials: Test of Speaking Skills

The most relevant task to test a student's speaking skills in ESP courses is a **Presentation** followed by a **Question and Answer** session. However, in order to guarantee an objective and fair assessment of a student's actual speaking skills, the following prerequisites need to be borne in mind.

Choice of Topic

Purely on the principle of authenticity, the choice of topic should be left to the individual student although caution is necessary in potentially sensitive areas such as politics, religion, etc. The following aspects should also be considered: How authentic is it to expect students to give a presentation on a topic they know little about? On the other hand, the choice should not simply be left to the students themselves, especially as from the B2 competence level upwards. Apart from the fact that the topic should be connected to the content or specialised area of the course, particularly in the case of ESP courses, the level of complexity both in terms of language and content required at the B2 and especially C1 levels can be extremely high if not impossible to achieve if the topic is too simplistic and as a consequence does not lend itself to the required degree of language complexity.

Structure

As the CEFR distinguishes two different speaking skills – *Spoken Production* (= monologue, e.g., presentations) and *Oral Interaction* (= dialogue, e.g., Questions and Answers), both speaking skills must be given equal weighting during the test. This requirement includes the need for the Questions to be as challenging as called for in the respective CEFR descriptors.

Context

The students should decide themselves the following features of the presentation they wish to give:

- the audience: academic, business, general audience
- the audience: presentation culture (especially as from the C1 level)
- the purpose: information, research, marketing

Individual or Team Presentation

Students occasionally express the wish to give a presentation as a team of two or more students. In my view and based on my teaching practice, such a format is not acceptable as it makes an objective view of the presentation and language skills of the individual students difficult, if not impossible.

Designing Language Test Materials: Test of Writing Skills

Probably the most relevant writing task for students in ESP or English for Academic Purposes (EAP) courses at the B2 level and higher is a Term Paper as that particular genre provides the opportunity to assess the students' ability to produce the style of language required in accordance with the descriptors for the respective level. An alternative task for the B2 level could be formal correspondence.

In order to guarantee an objective and fair assessment of a student's actual writing skills, the following prerequisites need to be borne in mind.

Choice of Topic

The choice of topic should be left to the student, but advice is often necessary, and caution is necessary in potentially sensitive areas such as politics, religion, etc. The following aspects should also be considered:

Complexity: Apart from the fact that the topic should be connected to the content or specialised area of the course, particularly in the case of ESP courses, it should also be a topic that lends itself to the level of complexity that is required at B2 and especially C1 level both in terms of language and content – a requirement that is often difficult to meet if the topic it too simplistic.

Length: The prescribed length is an important factor and needs to be given due consideration in connection with the topic and the degree of in-depth treatment the student is expected to provide. As a general rule, 2 – 2 ½ pages can be recommended as the minimum length for B2 and around 3 pages for C1. Details of size and font are given in the departmental guidelines.

Section 3: Portfolio-Based Assessment: Principles for Practice in an EFL Classroom

Tetyana Müller-Lyaskovets, Meni Syrou

Background

The origins of student portfolios as they are used today by German university language centers can be traced back to several sources. The European Language Portfolio (ELP) was first piloted and then launched in Europe in 2001 (Schneider & Lenz, 2001). In the US, portfolios were first introduced as a means of writing instruction and assessment in the 1980s, and now they are extensively implemented for a variety of purposes. North American academic institutions can use portfolios as mandatory entry and exit assessments across institutional boundaries. North American writing programs rely on portfolios as a tool for teaching, learning, and evaluation of native and nonnative language users. In our Department of Foreign Languages, we implement portfolios as a means of instruction and task-based performance assessment at the B2 and C1 levels in English writing courses, German studies and history courses (Landeskunde), English for Engineering, and German for Engineering.

Definition

Depending on their use, purpose, and geopolitical context, student portfolios can be conceptualized in several ways. In the North American education tradition, portfolios are defined ***as "a purposeful collection of students' artifacts created over time to display their efforts, growth and achievements to themselves, teachers, parents and other key stakeholders"*** (Genesee & Upshur 1996, cited in Lam 2014). The purpose of the ELP is "to provide a record of the linguistic and cultural skills they [learners] have acquired (to be consulted, for example, when they are moving to a higher learning level or seeking employment at home or abroad)" (Schneider & Lenz, 2001, p. 3). Similar to course or entry- and exist-level portfolios in the North American tradition, the ELP is seen as a tool that sustains learners' motivation to develop their language skills (Schneider & Lenz, 2001, p. 3). Both European and North American educational traditions recognize student portfolios as a unique tool for language learning, teaching, assessment, and fostering independent learning. Whereas at North American universities one use of a portfolio is to provide a basis for summative assessment at the end of a course, the ELP is not a course portfolio or a substitute for a language test (Schneider & Lenz, 2001, p. 6). The use of the ELP goes beyond the boundaries of a separate language course to be a document facilitating better mobility and educational exchange in Europe.

Problem Statement

Numerous publications have stressed that language learning is a process. In Europe, this emphasis on the process of language learning developed through a number of important publications such as, for example, the 1978 British Council Report *The Foreign Language Learning Process* or a more recent 2018 CEFR *Companion Volume With New Descriptors*.

Although there is a general consensus about language learning as **a multifaceted process**, a lot of German university language centers and departments still build their assessments in a way that prioritizes a single product. Thus, for example, students are asked to take a timed test or, in the case of writing, to write a timed single-draft essay. We view portfolio assessment (PA) to be one of the possible answers to this pedagogical predicament. At our department, we use portfolios as a teaching and assessment format that, first, capitalizes on the process nature of language learning and, second, includes this process in instruction and assessment of student performance. The authors of *The European Language Portfolio* convincingly argue that there are aspects of communicative competence, related to the *experience* of language learning, "which are not, or not necessarily, related to one specific level of language proficiency. These include learning strategies as well as socio-cultural and intercultural competence" (Schneider & Lenz, 2001, p. 49). Portfolios offer a format for the inclusion of these *experiences* in instruction and assessment.

Before embarking on a PA track, language instructors need to ask themselves a question: Can there be a high-quality product, such as, for example, an essay, a presentation, or a report, without a process of preparing this product? Does this process need to be taught? Does it deserve to be acknowledged through assessment? Does the process matter even if we deal, for example, with spontaneous speaking?

Why Portfolios?

In creating traditional language tests, language instructors at German universities are required to use the Common European Framework of References (CEFR) descriptors as a basis for their test development. *The European Language Portfolio* is a CEFR-based tool. However, its authors make a good point by showing that there are competencies that cannot "be linked to specific [CEFR] levels of competence consistently. Examples of such competencies are:

- sociocultural and intercultural competence including explicit knowledge about history, literature, art, the "way of life", etc. of the regions where a language is spoken;
- awareness of the variety of existing languages and their differences and common features;
- learning techniques and strategies" (Schneider & Lenz, 2001, p. 35).

Portfolio assessment is an effective alternative to more traditional forms of assessment of written and oral production because in addition to showcasing the learning process, it captures the multifaceted nature of language learning and production. It allows for a more genuine **performance assessment,** which is an authentic assessment of how students perform on a certain task or project rather than on answering a list of questions, no matter how authentic these questions are intended to be. As a performance assessment, PA gives more information than a traditional language test. For one thing, PA shows a student's performance growth during the course. For this reason, we believe that, in a classroom context, it is indispensable to students, instructors, and the institution.

The context in which language users communicate is complex, open, and variable. Think about the multiplicity of situations that require you to use language. Think about the multiplicity of genres that you use to communicate -- from writing an email, engaging in a casual conversation, giving a talk, participating in a Q&A to writing a report or an academic essay. Portfolios are a **flexible and open form of assessment** that can accommodate more genres and formats than end-of-course tests and that lends itself better to this open and variable context with a great variety of communication forms.

Finally, PA is a **more democratic** form of assessment. It involves students in the assessment process and by doing so makes them more responsible for and invested in their learning processes. By allowing students to exercise some control over their assessment, portfolios foster **self-regulated learning** that is known to have very high impact on student achievement. More direct involvement translates into learning more. As a democratic form of assessment, PA opens doors for a possible exchange, collaboration, and interdisciplinarity. It encourages students to see the language skills that they developed during the course as interconnected rather than as split into discrete reading, listening, speaking, and writing. An ability to facilitate the interconnectedness of competencies, contexts, and skills is one of the reasons why the Association of American Colleges and Universities lists e-portfolios among the high-impact educational practices: "Because collection over time is a key element of the ePortfolio process, employing ePortfolios in collaboration with other high-impact practices provides opportunities for students to make connections between various educational experiences" (Kuh & Schneider, 2008).

We believe that performance assessment in the form of course portfolios may be a more comprehensive, complete, and authentic indicator of student knowledge and skills than a traditional language test practiced by language centers and FL departments in Germany. In this context, a test is understood as "a collection of questions or tasks (which we generally call items), targeted to measure (provide quantifiable information about) learning objectives or some other trait" (Rodriguez & Albano, 2017, p. X). FL departments are not test centers, but instead, their mission is to be highly invested in teaching. Assessment should be viewed as an instrument of this teaching rather than its final purpose. Those who choose to practice PA believe that assessment has to be integrated with the curriculum as an extension of learning. Portfolios lend themselves to this integration.

Answering the question "Why portfolios now?" twenty years ago, Hamp-Lyons and Condon (2000) wrote, "… because portfolios answer today's need for a measurement system that can have a generative, rather than a reductive effect on education, because portfolios reinforce what we know about good teaching practice, because portfolios help teachers help learners assume more responsibility for their own learning, and because portfolios provide a rich source of information to teachers … and to researchers and administrators as they continue to assess educational progress ..." (p. XV).

In the context of teaching writing, research shows that PA is beneficial for "nonmainstream writers," that is mainly the EFL and ESL writers, because portfolios "provide a broader measure of what students can do" and allow for more time and revision to create a product (Hamp-Lyons & Condon, 2000, p. 61). In a PA classroom, time-pressure cannot be used as an excuse for poor work.

Characteristics of Course Portfolios

Course portfolios are usually defined through a number of characteristics such as collection, range, context richness, delayed evaluation, selection, student-centeredness, self-reflection, self-evaluation, and growth over time (Hamp-Lyons & Condon, 2000). In the following, we elaborate on each characteristic, suggested by Hamp-Lyons and Condon (2000), as it applies to our teaching practices at the Department of Foreign Languages, TU Dortmund University.

- **Collection**. Portfolios are collections of texts written in different genres. These texts display different competencies. A course portfolio usually opens with *a portfolio cover letter* (a reflective statement about the purpose of the portfolio and the author's literacy experiences), contains *major projects of the course*, and ends with a student *self-reflection and self-assessment*. Hamp-Lyons & Condon (2000, p. 30) argue that the collection feature contributes to the portfolio's higher face validity in the sense that students see portfolios as transparent and relevant to what is being assessed.

- **Range.** Portfolios have a high face validity when they require students to collect the artifacts of their *different performances and tasks* within a course. When assessing students on their portfolios, we base our judgments on a "range of performances" (Hamp-Lyons & Condon, 2000, p. 33). Please see our portfolio checklists on pages 33 – 37 for possible tasks and performances that we use.

- **Context richness**. This feature is usually viewed as one of the portfolio's biggest strengths. The work that students include in their portfolios results not from a situation of an exam but rather is an integral part of their learning: "This characteristic means that *instruction and assessment are intertwined*, that the context within which the learning took place determines the contents of the portfolio" (Hamp-Lyons & Condon, 2000, p. 34). For this feature to work, we need to teach the course as a succession of small but doable steps, each taking our students closer to performing on a larger task, such as, for example, giving a talk or writing a paper or a collection of essays. A course portfolio may reflect both student progress and their performance on major tasks.

- **Delayed evaluation**. Although students work on their artifacts all semester, they collect and submit them as a portfolio at the end of the semester. By delaying evaluation, we encourage students *to revise their work and exercise some control over their success*. To teach students how to revise their work, we should offer feedback on their work in progress. By doing so, we also teach them how to work on their own learning processes and how to reflect on their work. The task of writing the final self-assessment is easier for them if they had an opportunity to reflect on their work in progress at earlier stages. In terms of teaching, delayed evaluation can offer teachers some valuable information about the effectiveness of their courses (Hamp-Lyons & Condon, 2000, p. 34).

- **Selection**. Usually, students produce a larger corpus of work than they select for their portfolio. By choosing which works to submit, students have some control over self-representation (Hamp-Lyons & Condon, 2000, p. 35). When selecting artifacts for their portfolios, students have to make decisions about quality, which means that they practice self-assessment.

- **Student-centeredness**. Student-centeredness puts learner responsibility in the center of a learning experience. Students can decide what and how they study. By allowing students to revise and select their work, we give them some control over their success. If students receive early feedback, they can revise, and then “effort and time on task are explicitly rewarded” (Hamp-Lyons & Condon, 2000, p. 35).

- **Self-assessment**. Self-assessment is linked to self-regulated learning. By discussing their learning processes that led to their success, students draw implications for their future learning. Self-assessment promotes continuous learning.

- **Growth over time.** Portfolios may display student growth along specific parameters over time. *Process portfolios* may contain successive drafts or unrevised work, and *product portfolios* may contain earlier and later work, all of which allow for the evaluation of a degree of progress. For this feature to work, instructors have to organize their courses to reflect the process of learning. Instructors should also make assessment criteria available to students (Hamp-Lyons & Condon, 2000, pp. 36-37).

In designing their course portfolios, instructors can decide which portfolio characteristics they want to make more prominent depending on their course objectives and teaching approaches. These decisions define portfolio design. The following classification draws on Hamp-Lyons and Condon’s (2000) description of theories of writing but can be adapted to describe the types of language course portfolios.

Types of Course Portfolios

Based on course objectives and approaches to teaching, your course portfolios can draw on the following design patterns:

- **Product portfolios.** *Formalist approach* requires students to learn from and emulate high-quality texts. With this approach, a portfolio becomes a selection of products in specific genres (Hamp-Lyons & Condon, 2000, p. 39).

- **Process portfolios.** *Expressivist approach and process pedagogy* prompt students “to see their development … as centered in themselves, rather than centered in the teacher or the class” (Hamp-Lyons & Condon, 2000, p. 49). When used in a writing classroom, the process approach focuses on writing behavior (Hamp-Lyons & Condon, 2000, p. 48). Not only writing classes can be process-based. Process pedagogy can be transferred to other activities, such as, for example, listening comprehension, presentations, or vocabulary work. Language acquisition is a process that has to be taught. Assignments may include keeping a journal that documents student growth and extensive drafting.

- **Reader-centered portfolios**. Cognitive theories focus on how authors can incorporate *reader-friendly features* into their texts (Hamp-Lyons & Condon, 2000, pp. 42-43). We use language with a certain communicative purpose. A reader-centered approach asks students to make informed choices about how they communicate (organize and present) information in a way most accessible to readers. What do we take from this approach? When teaching and assessing, we take into consideration “how easily readers can process and remember the texts in a … portfolio" (Hamp-Lyons & Condon, 2000, p. 43). In addition to major projects, reader-centered portfolios might include the following assignments: a presentation review, paper review, documentary or film review, or author-purpose-means analysis of a written or oral text. Reader-centered portfolios work well for teaching technical communication courses and presentation skills courses.

- **Collaborative learning portfolios.** As Hamp-Lyons and Condon (2000) note, "collaborative learning assumes that writing and knowledge develop among individuals, not in the individual" (p. 44). We learn better when we collaborate and share our learning with each other. Groups can work towards a portfolio assessment, and peers can "share their developing portfolios at regular intervals" (Hamp-Lyons & Condon, 2000, p. 45). This type of portfolio works well also for group projects.
- **Discourse community portfolios.** Discourse community is a group of people who share certain discourses and communication mechanisms among themselves. For example, writers within a certain discipline form a discourse community. The authors of this portfolio type should ask themselves a question: How do "members of a particular discourse community create and convey knowledge?" (Hamp-Lyons & Condon, 2000, p. 46). This portfolio type works well for teaching English within a disciple or ESP courses. The purpose of a discourse community portfolio will be to ask students to use forms, strategies, and skills that a student needs to become a member of a particular discourse community (Hamp-Lyons & Condon, 2000, p. 46). In addition, a classroom based on this pedagogy will foster extensive interaction among its participants/members of a discourse community.

Challenges of Portfolio Assessment

The use of portfolios usually stems from a certain pedagogy practiced by an institution or a program. If the practice of PA is consistent with this dominant pedagogy, then we can make consistent judgments about student performance in a PA classroom. However, this successful merger is seldom the case. Thus, for example, the 2018 CEFR *Companion Volume with New Descriptors* argues for the *action-oriented rather than four-skill-oriented approach*. *The Companion* offers a different model of communicative activities that now include *reception, production, interaction, and mediation*. However, a lot of language centers and departments continue to conceptualize their curriculum and assessment by referring to the four skills of reading, listening, speaking, and writing. Under this circumstance, the question is: How do we implement the four-skill assessment (25% per skill) format in a portfolio classroom?

Another challenge is the degree of formalization and codification practiced by some language centers. To facilitate the consistency of practice across all CEFR-informed assessments, quite often language departments require their teachers to use uniform grading rubric for tasks and skills across all language courses of a certain level. An example of such a rubric is a grading rubric for presentations at the B2 level that is used to grade student presentations in a four-skill course, such as, for example, B2 English for Engineering, and in a presentation skills course. Hamp-Lyons and Condon (2000) argue that "portfolio assessment cannot be fully effective in a top-down model" (p. 64). Teachers and even students (all stakeholders) should be involved in the design of portfolios and corresponding grading rubrics. In terms of assessment, the department can provide a shared rubric for portfolios. However, the instructors should be free to customize this document and tailor it to the needs of their courses.

Costs, or time spent on grading, are usually mentioned among the drawbacks of PA. However, developing a new high-quality language test each semester is a no less demanding and time-consuming task. Our experience of teaching PA courses shows that providing feedback on work in progress expedites grading of the final draft. In addition, grading portfolios is a rewarding experience that allows for a glimpse into what worked well and what did not work for your students in your class.

Finally, introducing PA in your teaching is a step that may mark a change in the underlying pedagogy. Hamp-Lyons and Condon (2000) argue that “portfolio is a change agent” (p. 64). One should be aware of this quality when switching to PA. PA will expose our weaknesses in teaching (the lack of learning, for example) and assessment sooner and in a more powerful way than an end-of semester test.

Portfolio Classroom Practices and Assignments

What Kind of Tasks and Assignments Can Facilitate Learning in a PA Classroom?

- Ask your students to create multiple drafts of a paper. Multiple drafting will not only improve their language control but also help them learn more about their own ideas.
- When asking your students to revise, give them a specific direction. Revising for all possible flaws at a time is difficult. For example, you can ask them to revise a paper for structural coherence. Next, move to the level of a paragraph to see if paragraphs are tight. At the next step, ask them to check their vocabulary. Finally, require your students to proofread their drafts for mechanics. Create checklists to facilitate revision.
- Organize your course as a succession of small but doable steps and ask your students to include their work on different steps. For example, when preparing a presentation, you may ask them (1) to write a topic proposal and discuss it with their peers, (2) create a presentation outline, (3) create the slides, and (4) create a script. Finally, after the presentation, ask them to write a review of one of their peer’s presentations.
- Do a writer’s workshop: a student reads out a paper aloud and the class discusses it. This method shows a writer or a speaker how their message is perceived by the audience.
- Extend the peer critique assignment by asking a writer or a speaker to write a reflection on peer feedback.
- Have students engage in collaborative learning. You can use a learning community approach where you group students and have them engage in peer reviewing of each other’s work during the entire semester.

- Practice reflective writing such as keeping a journal and writing portfolio cover letters and self-assessments.

Final Comments

Before introducing portfolios into your classroom, reflect on your teaching pedagogy and values. As the next step, design your portfolio and create assessment criteria. Make these criteria available to students at the beginning of the course. You may also involve students in creating PA criteria, which will allow them to reflect deeper on their learning practices and outcomes. Collaborate with your colleagues and be prepared to be challenged. Reflect on the generative effect that portfolios have on motivation, learning, and knowledge-making. Have you ever heard anyone saying, “I wish I could take one more test or write one more exam?” Yet, in many ways, we owe the creation of new knowledge to people, such as scientists and scholars among others, who find writing papers, working on projects, and giving talks to be rather rewarding activities. Portfolios allow for engaging your students in the above-mentioned activities and genres, which your students might be dealing with in their future careers.

Resources for Teachers

English Courses

***Handout 1* – C1 Academic Writing and Reading**

Portfolio Checklist

- ✓ Cover page
- ✓ Portfolio checklist
- ✓ Portfolio cover letter
- ✓ One-page paper proposal
- ✓ Paper outline (this is NOT a table of contents)
- ✓ First draft of your essay (you can print it out with my comments)
- ✓ Revised copy of a **seven-page/ten-page essay: 2800 words**
- ✓ One-page author-purpose-means analysis (graded). If you want to revise it, please attach the original graded copy and the improved draft.
- ✓ Vocabulary records
- ✓ Half-page self-evaluation
- ✓ Grading rubric

***Handout 2* – B2 Writing**

Portfolio Checklist

- ✓ Cover page
- ✓ Portfolio checklist
- ✓ Half-page introduction that explains the purpose of the portfolio
- ✓ Half-page paper proposal

- ✓ Paper outline (this is NOT a table of contents)
- ✓ First draft of your essay (you can print it out with my comments)
- ✓ Revised copy of a **three-page/five-page essay: 1500 to 1800 words**
- ✓ One-page response essay (graded). If you want to revise it, please attach the original graded copy and the improved draft.
- ✓ Vocabulary records
- ✓ Half-page self-evaluation
- ✓ Grading rubric

***Handout* 3 – C1 Presenting and Discussing Ideas**

Portfolio Checklist

- ✓ Portfolio checklist
- ✓ Introduction. In one or two paragraphs, tell us about the purpose of your portfolio and what achievements it illustrates.
- ✓ Outline of your two-minute story performance. See a separate handout for this activity.
- ✓ Outline of your two-minute performance “Explaining a Concept/Process.” See a separate handout for this activity.
- ✓ Message box for your ten-minute presentation
- ✓ Ten-minute presentation topic proposal (one double spaced page, font 12)
- ✓ Outline of your ten-minute presentation
- ✓ One-page review of your peer’s presentation (double spaced, font 12). Use the form/template on Moodle.
- ✓ Three journal entries for your peers’ presentations. Use the form/template on Moodle.
- ✓ Your ten-minute presentation slides. You can put several slides on one page.
- ✓ Vocabulary records for three course readings and five videos of your choice. Use the form/template on Moodle.
- ✓ Conclusion. In one paragraph, reflect on your learning outcomes for the class and on the challenges that you dealt with in this class.

***Handout 4* – B2 English for Engineering**

Portfolio Checklist

- ✓ Cover page (please use the template)
- ✓ Portfolio checklist (you can use this document)
- ✓ Portfolio cover letter that explains the purpose of your portfolio (the max. length is one page)
- ✓ One-page presentation proposal (use the message box)
- ✓ Presentation outline (graded as work in progress)
- ✓ One-page presentation review
- ✓ Three short essays of your choice that demonstrate your writing skills (1500 to 1800 words total)
- ✓ Documentary review (400 to 500 words)
- ✓ Half-page self-evaluation
- ✓ Grading rubric

German as a Foreign Language Courses

Handout 5 **– Deutsch B2/C1 Landeskunde**
Portfolio Checklist ✓ Cover page (template on Moodle) ✓ Portfolio checklist (template on Moodle) ✓ Portfolio cover letter that explains the purpose of the course from your point of view as well as the purpose of your portfolio (one page in accordance with the template) ✓ Self-assessment and goals for the course (refer to CEFR self-assessment rubric for input) ✓ One-page presentation proposal (template and guidelines on Moodle) ✓ Presentation outline (graded as work in progress) ✓ Presentation review (one page in accordance with the template) ✓ Selection of your written assignments in the revised version (please choose 4 texts) ✓ Two further written tasks (topic and text-type to be chosen independently on the basis of the overall content of the course) ✓ Exemplary demonstration of independent vocabulary work (use input and handouts on Moodle) ✓ One page self-evaluation (refer to the CEFR self-assessment rubric for input)

Handout 6 **– Fachsprache Deutsch B2/C1 für Ingenieurwissenschaften**
Portfolio Checklist ✓ Cover page (template on Moodle) ✓ Portfolio checklist (template on Moodle) ✓ Portfolio cover letter that explains the purpose of the course from your point of view as well as the purpose of your portfolio (one page in accordance with the template) ✓ Self-assessment and goals for the course (refer to the CEFR self-assessment rubric for input) ✓ One-page presentation proposal (guidelines on Moodle) ✓ Presentation outline (graded as work in progress) ✓ Presentation review (one page in accordance with the template) ✓ Selection of your written assignments in the revised version (please choose 4 texts) ✓ Two further written tasks (topic and text-type to be chosen independently on the basis of the overall content of the course) ✓ Exemplary demonstration of independent vocabulary work (use input and handouts on Moodle) ✓ One page self-evaluation (refer to the CEFR self-assessment rubric for input)

Table 2. Grading Rubric. Portfolios at C1 Level Academic Writing

Portfolios at C1 Level Academic Writing	Very Good (5 Points)	Good (4 Points)	Insufficient (3 Points)	Poor/ Inadequate (0 Points)	**Comments**
Progress, Clarity, Organization, Focus, Flow, Evidence, Critical Thinking, and Research Techniques					
The writer effectively **documents his/her learning experiences** related to course learning outcomes.					
The portfolio provides evidence of **self-evaluation and reflection** to examine what produced the candidate's growth as a writer.					
The writer uses **revision** to improve the quality of writing.					
The portfolio **layout** is professional. All sections are clearly titled. All artifacts are clearly labeled.					
All papers are **readable immediately throughout**.					
Clarity: Vocabulary and sentence structure are sophisticated and appropriate for the topic and intended audience. Published documents enhance the reader's understanding of the topic and allow for **a smooth reading experience**.					
Organization/Flow: At the **level of a paper**, the writer (1) organizes ideas clearly and consistently and (2) maintains consistent focus and sense of purpose.					
Organization: The writer effectively structures and orders **paragraphs** and links ideas with smooth **transitions**.					
Focus: Main points are clearly stated and well advanced.					**Date of Grading:**
Evidence: There is enough evidence to support main points.					
Critical Thinking: The writer (1) engages in an insightful, well-articulated analysis, synthesis, and critique of the subject and (2) thoughtfully considers multiple positions where appropriate.					
Research Techniques: The writer (1) skillfully uses appropriate materials from a variety of sources to support ideas and (2) consistently uses a standard documentation style.					**Examiner's Signature:**
Mechanics and Usage					
Writing is virtually free from the kinds of errors that distract from meaning and readability.					
All papers are free from redundancies. All papers have been revised for **wordiness**.					
The **grammar** is correct throughout.					
A variety of different grammatical constructions is used.					
The writer uses correct, sophisticated, and varied sentence structures.					
The **vocabulary** is correct in usage and appropriate to the target audience.					
A broad spectrum of vocabulary (e.g., synonyms, descriptions, definitions – not translations) is used.					
The **spelling** and **punctuation** are correct throughout.					
Total Score:				**/100**	
Pass Grades (please circle): 1,0 (95.5-100) / 1,3 (91.5-95) / 1,7 (87.5-91) / 2,0 (83.5-87) / 2,3 (79.5-83) / 2,7 (75.5-79) / 3,0 (71.5-75) / 3,3 (67.5-71) / 3,7 (63.5-67) / 4,0 (60-63) **Fail Grades (please circle): 5,0 (32-59) / 6,0 (0-31)**					

Table 3. Grading Rubric. Portfolios in the German as a Foreign Language Classroom at B2/C1 Level

Portfolios – Deutsch B2/C1 Landeskunde und Deutsch B2/C1 Fachsprache für Ingenieurwissenschaften	Sehr gut (5 Punkte)	Gut (4 Punkte)	Ausreichend (3 Points)	Mangelfhaft/ Ungenügend (0 Points)	**Kommentare:**
Textbezogene Aspekte (Aufbau, Fokus, Leserfreundlichkeit, kritisches Denken, wissenschaftliche Arbeitstechniken)					
Der Verfasser bzw. die Verfasserin des Portfolios dokumentiert effektiv seine/ihre Lernerfahrungen im Zusammenhang mit den Lernzielen des Kurses.					
Das Portfolio stellt die Fähigkeit zur Reflexion über die eigenen Lernerfahrungen und den eigenen Lernerfolg mit Blick auf die Lernziele des Kurses unter Beweis.					
Die Portfolio-Beiträge wurden überarbeitet. Textfeedback wurde berücksichtigt.					
Das Portfolio Layout ist professionell gestaltet. Alle Kapitel/Abschnitte sind anschaulich betitelt. Alle Portfolio-Beiträge sind anschaulich gekennzeichnet.					
Die Portfolio-Beiträge sind durchgängig leserfreundlich gestaltet und ermöglichen ein flüssiges Leseerlebnis.					
Der Wortschatz und Satzbau der Portfolio-Texte sind anspruchsvoll und dem Thema und dem Zielpublikum angemessen. Die im Portfolio dargelegten Beiträge verbessern das Verständnis des Lesers für das übergreifende Thema.					
Die Struktur der Portfolio-Beiträge ist logisch und kohärent; die Paragraphen werden sinnvoll eingesetzt.					
Die Argumentation der Portfolio-Beiträge ist schlüssig und klar formuliert.					**Datum der Bewertung:**
Die Portfolio-Beiträge enthalten, dort wo angebracht ausreichend Belege, um die Hauptpunkte zu unterstützen.					
Der Verfasser bzw. die Verfasserin des Portfolios (1) führt eine aufschlussreiche, gut artikulierte Analyse, Synthese und Kritik des Themas durch und (2) zieht, wo es angebracht ist, mehrere Positionen in Betracht.					**Unterschrift Prüfer*in:**
Der Verfasser bzw. die Verfasserin des Portfolios (1) verwendet geeignete Materialien aus einer Vielzahl von Quellen, um Ideen zu unterstützen und (2) verwendet konsequent einen Standard-Dokumentationsstil.					
Die vertiefenden Portfolio-Beiträge zu selbstgewählten Themenstellungen stellen ein kritisches und eigenständiges Denken unter Beweis. Die Textsorte ist angemessen gewählt.					
Sprache (Grammatik, Wortschatz, Interpunktion)					
Die Portfolio-Texte sind insgesamt frei von solchen sprachlichen Fehlern, welche die Lesbarkeit beinträchtigen.					
Die Portfolio-Texte sind frei von sprachlichen Redundanzen und Füllwörtern.					
Die Grammatik ist weitgehend korrekt.					
Es wird eine Bandbreite an unterschiedlichen grammatikalischen Konstruktionen eingesetzt.					
Die Portfolio-Texte weisen einen korrekten, differenzierten und abwechslungsreichen Satzbau auf.					
Der verwendete Wortschatz ist korrekt und dem Zielpublikum angemessen.					
Der verwendete Wortschatz verfügt über eine Bandbreite an Begriffen, z.B. an Synonymen, Definitionen.					
Die Rechtschreibung und Interpunktion sind durchgängig korrekt.					
Gesamtpunkte:				**/100**	
Bestanden (bitte ankreuzen): 1,0 (95.5-100) / 1,3 (91.5-95) / 1,7 (87.5-91) / 2,0 (83.5-87) / 2,3 (79.5-83) / 2,7 (75.5-79) / 3,0 (71.5-75) / 3,3 (67.5-71) / 3,7 (63.5-67) / 4,0 (60-63) **Nicht bestanden (bitte ankreuzen): 5,0 (32-59) / 6,0 (0-31)**					

References

The Common European Framework of Reference for languages: Companion volume with new descriptors. (2018). Council of Europe. https://icc-languages.eu/common-european-framework-of-reference-for-languages-learning-teaching-assessment-companion-volume

Kuh, G. D., & Schneider, C. G. (2008). *High-impact educational practices: What they are, who has access to them, and why they matter*. Association of American Colleges and Universities. https://www.aacu.org/node/4084

Lam, R. (2014). Promoting self-regulated learning through portfolio assessment: Testimony and recommendations. *Assessment & Evaluation in Higher Education*, *39*(6), 699-714. https://doi.org/10.1080/02602938.2013.862211

Rodriguez, M. C., & Albano, A. D. (2017). *The college instructor's guide to writing test items: Measuring student learning*. Routledge.

Schneider, G., & Lenz, P. (2004). *The European Language Portfolio*. Lern- und Forschungszentrum Fremdsprachen, University of Fribourg/CH. https://rm.coe.int/1680459fa3

The Foreign Language Learning Process. (1978). The British Council. https://www.teachingenglish.org.uk/article/foreign-language-learning-process

Section 4: Using Benchmarks to Assess Student English Language Proficiency Levels in the DAAD Exams

Kai Herklotz, Tetyana Müller-Lyaskovets

The DAAD Language Certificate is intended to be a proficiency rather than achievement assessment. It attests a student's current language level and is oftentimes required as part of an application for short-term study abroad, internship abroad, scholarships, or exchange programs. Universities, programs, or companies may accept or even require other standardized tests such as TOEFL (Test of English as a Foreign Language), Cambridge Certificate of Advanced English, or IELTS (International English Language Testing System). Yet, the DAAD Language Certificate is oftentimes faster, less complicated, and less expensive for students.

Only TU Dortmund University students can take the DAAD exam at our Department of Foreign Languages. Students are not allowed to take the exam more than once in six months. There is an option for students to earn partial or full DAAD test credit on the basis of courses the student has already successfully completed at the Department of Foreign Languages within the last two semesters.

The examination required for the issue of the DAAD Language Certificate comprises a written and an oral part and tests the existing knowledge and skills in the areas of reading comprehension, written production, and spoken interaction.

The written part of the examination assesses written expression via a 200- to 250-word expository essay. Students are given a choice of three to four topics, and the time limit to complete the writing assignment is 30 minutes. No dictionaries are allowed. The students also submit a letter of motivation as a take-home written task. The letter of motivation is used to assess writing skills. In addition, it serves as a teacher prompt for the oral part of the exam.

The oral part of the examination assesses reading comprehension and spoken interaction as an integrated skills task. It consists of two parts: 1) questions about a 750- to 1000-word reading text tailored to the language level and 2) interview-style questions about the student's educational background, university degree program, internships and other relevant educational and practical experience, and career goals. More detailed information about the test format is available on the departmental website.

The DAAD Language Certificate covers the range from B1 to C1 of the CEFR, and all examination tasks have to be chosen and developed accordingly. Writing prompts have to allow for answers ranging in complexity and depth. Readings will vary slightly in length and complexity according to the student's desired language level. During spoken interaction, the instructor adjusts their delivery to accommodate the student's ability to follow and contribute to the conversation to a degree that shows the student's language level.

We evaluate the written production part of the DAAD examination according to the CEFR descriptor scales for "Written reports and essays" (Council of Europe. Education Department. Education Policy Division. Language Policy Programme, 2018, p. 76).

We need to differentiate between giving impressions and opinions (A2), giving and justifying an opinion (B1), giving reasons in support of or against a particular point of view (B2), and expanding and supporting points of view with subsidiary points, reasons, and relevant examples (C1). Additionally, we need to differentiate between listing (B1) and explaining (B2) advantages and disadvantages, and producing more complex texts at the C1 level.

We evaluate the reading comprehension part of the DAAD examination according to the CEFR descriptor scales for "Reading for information/argument" (Council of Europe. Education Department. Education Policy Division. Language Policy Programme, 2018, p. 63).

We assess depth of understanding, ranging from understanding points of view and main conclusions in texts on subjects of current interest (B1), to obtaining information, ideas, and opinions (B2), to identifying finer points of detail including attitudes and implied as well as stated opinions (C1).

We evaluate the spoken interaction part of the DAAD examination according to the CEFR descriptor scales for "Spoken interaction," especially "Understanding an interlocutor" and "Formal discussion," as well as for "Production strategies" such as "Compensating" and "Monitoring and repair" (Council of Europe. Education Department. Education Policy Division. Language Policy Programme, 2018, pp. 83-87). We assess understanding in detail of complex topics (C1), understanding in detail (B2), and understanding in everyday conversation (B1).

We also assess clear, detailed, smoothly flowing descriptions of complex subjects and abstract topics (C1), systematically developed descriptions with relevant supporting detail and relevant examples (B2), and straightforward description as a linear sequence of points with some accommodation by the interlocutor (B1).

References

Council of Europe. Education Department. Education Policy Division. Language Policy Programme. (2018). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Companion Volume with New Descriptors*. Council of Europe. https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989

Example of DAAD Rubric: Speaking

Task: (1) Discuss the journal/magazine article you've just read. (2) Answer your instructor's questions about the ideas in your letter of motivation and explain the goals for your study abroad.

Table 4. TOEFL iBT Test Independent Speaking Rubrics

Level	General Description	Delivery	Language Use	Topic Development
C1	The response fulfills the demands of the task, with at most minor lapses in completeness. It is highly intelligible and exhibits sustained, coherent discourse. A response at this level is characterized by all of the following:	Generally well-paced flow (fluid expression). Speech is clear. It may include minor lapses, or minor difficulties with pronunciation or intonation patterns, which do not affect overall intelligibility.	The response demonstrates effective use of grammar and vocabulary. It exhibits a fairly high degree of automaticity with good control of basic and complex structures (as appropriate). Some minor (or systematic) errors are noticeable but do not obscure meaning.	The response is sustained and sufficient to the task. It is generally well developed and coherent; relationships between ideas are clear (or clear progression of ideas).
B2	The response addresses the task appropriately but may fall short of being fully developed. It is generally intelligible and coherent, with some fluidity of expression, though it exhibits some noticeable lapses in the expression of ideas. A response at this level is characterized by at least two of the following:	Speech is generally clear, with some fluidity of expression, though minor difficulties with pronunciation, intonation, or pacing are noticeable and may require listener effort at times (though overall intelligibility is not significantly affected).	The response demonstrates fairly automatic and effective use of grammar and vocabulary and fairly coherent expression of relevant ideas. The response may exhibit some imprecise or inaccurate use of vocabulary or grammatical structures or be somewhat limited in the range of structures used. This may affect overall fluency, but it does not seriously interfere with the communication of the message.	The response is mostly coherent and sustained and conveys relevant ideas/information. The overall development is somewhat limited, usually lacks elaboration or specificity. Relationships between ideas may at times not be immediately clear.

B1	The response addresses the task, but development of the topic is limited. It contains intelligible speech, although problems with delivery and/ or overall coherence occur; meaning may be obscured in places. A response at this level is characterized by at least two of the following:	Speech is basically intelligible, though listener effort is needed because of unclear articulation, awkward intonation, or choppy rhythm/pace; meaning may be obscured in places.	The response demonstrates limited range and control of grammar and vocabulary. These limitations often prevent full expression of ideas. For the most part, only basic sentence structures are used successfully and spoken with fluidity. Structures and vocabulary may express mainly simple (short) and/or general propositions, with simple or unclear connections made among them (serial listing, conjunction, juxtaposition).	The response is connected to the task, though the number of ideas presented or the development of ideas is limited. Mostly basic ideas are expressed with limited elaboration (details and support). At times relevant substance may be vaguely expressed or repetitious. Connections of ideas may be unclear.

Proficiency Level Benchmarks

C1 Level Speaking

The student provides a *well-supported* summary of the article by identifying the key ideas and by supporting them with the evidence from the text. The student can express and support their opinion about ideas found in the text. Minor lapses can occur, but the message is clear, *well-articulated*, and clearly delivered. The student answers interview questions clearly, elaborates on his or her statements, avoids unsubstantiated generalizations, and speaks with a certain degree of *flair*.

B2 Level Speaking

The student provides a *mostly complete* summary of the article by identifying the key ideas and by supporting them with the evidence from the text. Yet, some ideas may be omitted, not fully developed, or lack elaboration. The student can express and support their opinion about the ideas found in the text but may experience difficulties with some demanding content. Minor lapses can occur, but the message is clear, well-articulated, and *in general* clearly delivered. Although the student answers interview questions clearly and elaborates on statements, his or her speech may be wanting in sophistication.

B1 Level Speaking

The student conveys *some main points* and other relevant information about ideas and facts from the article. Yet, the summary, explanations, and opinions are *sometimes* incomplete, inaccurate, and/or lack detail; long or complex explanations may lack coherence. The student *can speak clearly enough and without much hesitance about familiar topics, including his or her study abroad program*, but limitations in the range of vocabulary, grammar, and pronunciation may obscure the message conveyed when speaking spontaneously.

Source: ***TOEFL iBT***® Test Independent **Speaking** Rubrics with further adaptation

Section 5: Possible Summative Assessment Tasks and Activities at the B1 to C2 Levels of Proficiency

Tetyana Müller-Lyaskovets

The chapter delineates both *in-class* assessments (Table 5) and *take-home* projects (Table 6) used by our instructors to assess their students' achievement and performance at a certain CEFR-specific level of proficiency.

In this context, tests are understood as "a collection of questions or tasks (which we generally call items), targeted to measure (provide quantifiable information about) learning objectives or some other trait" (Rodriguez & Albano, 2017, p. X). Table 5 also lists other in-class assessments, such as, for example, single-draft in-class essays.

In-Class Assessments

Table 5. Exam Questions and Tasks at B1 to C1 Levels

Question Type	Skills	Levels
Note-taking, or open-ended questions Note-taking is the practice of recording information usually during lectures or from books to help one identify the important ideas from that source. Taking notes while reading or listening is an important skill one needs to succeed academically and professionally. Examples of questions that require note-taking: • "What has Black failed to do (mention two things)?" • "According to the passage, one of the journalist's worries about her present situation is that she..." Possible types of open-ended questions: • *Understanding the gist questions*: "What is the lecture mainly about?" • *Understanding the details, or little picture questions*: "According to the narrator, which three devices are relatively new to customers?" Ask about details that are *vital* to the understanding of the passage. • *Understanding attitude questions*: "How does the speaker feel about enforcing diversity?" • *Understanding connections questions*: "What example does Black provide in … What does this example illustrate?" (may also include *make comparisons questions* or *understanding cause-effect relationships questions*). • *Summary questions*: "Summarize the reasons White provides for …." • *Note-taking inference questions:* "What can you conclude about…? Based on the information in the passage, what can be inferred about…?" (These questions do not have a direct answer within the text, but rather answers can be inferred from the clues within the text.)	RC and LC	B1 to C2

Yes because/No because questions/True & False items with justification for choosing True or False Example: "Decide if the statements are true or not true according to the information in the text. Then give evidence from the text to support your answer. You will receive one point for each correct answer with evidence from the text." Because this type of task tests RC rather than text production, students can and should use sentences or phrases from the text. Grammar and spelling are not graded. One should use these items with caution to test LC because answering them requires not only listening comprehension skills but involves other cognitive processes.	RC and LC	B1 to C2
Text production Students are given a choice of **4 contextual topics in different genres** (for example, a blog entry, a wiki entry, an email, an answer for the frequently asked questions page, or a paragraph for a college essay) related to the audio or reading text or topics dealt with in class. Students write a 200- to 250-word text on one of the topics.	single draft in-class writing	B1 + B2
Report writing (350 – 400 words) based on an approximately ten-minute audio/video The report is graded for content, coherence, structure, accuracy and repertoire of grammar and vocabulary, orthography, paragraphing, and the appropriateness of register and style.	single draft in-class writing	C2

Take-Home Extended Tasks and Projects

Projects are especially suitable for **practicing and assessing production skills**. Besides, they require students to use higher-order thinking skills, such as problem-solving and critical thinking. Higher-order thinking skills are not level specific. Normally, these projects are described in terms of **task-based performance assessments** that evaluate student performance in real-life contexts.

Table 6. Take-Home Tasks and Projects

Project	**Skills**	**Level**
Portfolio Portfolios are defined ***as "a purposeful collection of students' artifacts created over time to display their efforts, growth and achievements to themselves, teachers, parents and other key stakeholders"*** (Genesee & Upshur 1996, cited in Lam 2014).	writing	B2 to C2
Presentation + Q&A session Depending on the course and student proficiency level, instructors can use different presentation tasks. Examples may include: telling a story, explaining a concept or a process, or giving a talk on a topic related to one's field of study or prospective employment.	speaking; writing (slides, outlines, and scripted talks)	B1 to C2

Facilitation Depending on the course and student proficiency level, instructors can ask students to facilitate a presentation or group discussion.	speaking	B2 to C2
Term paper	writing	B2 to C2
Annotated bibliography A bibliography is a list of sources that includes bibliographic information. An annotated bibliography also includes a summary and evaluation of each of the sources.	RC and writing	B2 to C2
Documentary review The goal of this assignment is to familiarize students with and assess the listening comprehension process required for a critical analysis of a text.	LC and writing	B2 to C2

You can find a detailed description of these tasks in the section Examination Guidelines on the English Teacher Moodle Platform.

Section 6: Tips and Strategies for Self-Editing Your Tests

Tetyana Müller-Lyaskovets

One of the most common complaints about the tests is “THE QUESTIONS ARE CONFUSING OR AMBIGUOUS” (*The College Instructor’s Guide to Writing Test Items*, 2017, p. XII). What can we do to achieve clarity in articulating the exam tasks and asking questions? These are some of the recommendations for *achieving clarity of expression* in item, question, and task writing. The recommendations are based on our review of the SoSe 2019 tests.

In no way being a comprehensive guide to item or task writing, these recommendations will draw your attention to some common defects when writing your items or explaining assignments to students.

Strategies and Tips for Clear and Explicit Question/Item Writing

- **Make sure that there are no interdependent items** (*The Manual for Language Test Development,* 2002, p. 18). Keep the content of each item independent of one another (*The College Instructor’s Guide*, p. 52). Make sure that each item stands on its own and makes a clear usage of pronouns.

  INCORRECT: In 1896, he founded an elementary school that would serve as a laboratory for the newly set up Department of Education. True/False
  CORRECT: In 1896, Black founded an elementary school that would serve as a laboratory for the newly set up Department of Education. True/False
  INCORRECT: It excludes the external stimulation of the sense organs, but not the response initiated by the internal activity of the child’s brain. True/False
  CORRECT: The concept of the reflex arc excludes the external stimulation of the sense organs, but not the response initiated by the internal activity of the child’s brain. True/False

- **Make sure that you are using the correct pronoun** so that the reader clearly understands which noun your pronoun refers back to.

  INCORRECT: Engaging your workforce to perform to their maximum potential, managers must learn how to fully utilize the talents of all their staff and not only the employees who are similar to themselves. True/False
  CORRECT: Engaging their workforce to perform to their maximum potential, managers must learn how to fully utilize the talents of all their staff and not only the employees who are similar to themselves.

  Resource: Using Pronouns Clearly. Purdue OWL:
  https://owl.purdue.edu/owl/general_writing/grammar/pronouns/index.html

- **Be direct and concise**. Minimize the amount of reading in each question/item. Avoid “window dressing.” (*The College Instructor’s Guide,* p. 52)

  WORDY: Because the telephone was invented, there were many new opportunities for better communication.

REVISED: The invention of the telephone created many opportunities for enhanced communication.

Source: Schleppegrell, p. 73, cited in Purdue OWL: https://owl.purdue.edu/owl/english_as_a_second_language/esl_students/tips_for_writing_in_north_american_colleges/concision.html

- **Avoid negative statements**. Negatively wording the statement creates a cognitively complex task (*The College Instructor's Guide*, p. 41, p. 126).

  CONFUSING/ CLUMSY: Difficulties in knowing how to turn ____________________into leadership practices may lead back to not possessing the skill of cultural fluency.
  REVISED: Difficulties in knowing how to turn ________________________into leadership practices may result in the lack of the cultural fluency skills.

- **Tap one aspect of content at a time**, asking only one question instead of two. Be careful using the word "and" that introduces multiple aspects of content. This usage complicates interpretation for the student (Which part do I answer?) and for the teacher (Which part was misunderstood?) (*The College Instructor's Guide,* p. 124)

  CONFUSING: What does Green tell us about architecture and what should be the role of architects in our society?
  REVISED: What does Green tell us about architecture? (XX points) What should be the role of the architects in our society? (XX points) / How does Green describe the role of the architects in our society? (XX points)
  CONFUSING: What type of data was used and what are the demands of the hackers? (XX points)
  REVISED: What type of data was used? (XX points) What are the demands of the hackers? (XX points)

- When you are tapping several aspects of content at a time, **provide explicit instructions as to how many answers your students have to provide.**

  EXAMPLE:
  What three details are given about the main person involved in the experiment? (three x one point)
  # ____
  # ____
  # ____

- **Edit and proofread your questions** for common structural defects:

  Run on sentences
  WYPR correspondent Emily White spoke with local residents about the cyberattack, what were their opinions?
  Resource: Fused sentences/run-on sentences
  https://owl.purdue.edu/owl/general_writing/punctuation/independent_and_dependent_clauses/runonsentences.html

Structures that are supposed to be parallel but are not parallel.
Resource: https://owl.purdue.edu/owl/general_writing/mechanics/parallel_structure.html

NOT PARALLEL:
Concerning the ransom, Baltimore Mayor Jack White has said the city
a) will pay the ransom within the month.
b) will not pay the ransom.
c) the State of Maryland will pay the ransom.
d) the NSA will replicate the key to unlock the data.
PARALLEL:
Concerning the ransom, Baltimore Mayor Jack White has said
a) the city will pay the ransom within the month.
b) the city will not pay the ransom.
c) the State of Maryland will pay the ransom.
d) the NSA will replicate the key to unlock the data.

Awkward placement of subordinate clauses and modifiers. In general, keep related words together.

CONFUSING:
Since 1990 which of the following has Erik Green founder of the "FUCT" clothing line said has cost him real money?
a) Counterfeiters
b) Producers of knockoffs
c) Copycats
d) All of the above
REVISED:
The founder of the "FUCT" clothing line Erik Green said that several factors have cost him real money since 1990. Which of the following are those factors?
a) Counterfeiters
b) Producers of knockoffs
c) Copycats
d) All of the above
Resource on sentence clarity:
https://owl.purdue.edu/owl/general_writing/mechanics/sentence_clarity.html

Avoid gender-biased language

ORIGINAL USE: Which symbolism did Black White, chairman of CCS Fundraising, use to describe Notre Dame Cathedral?
ALTERNATIVE USE: Which symbolism did Black White, chairperson of CCS Fundraising, use to describe Notre Dame Cathedral?
Resource:
https://owl.purdue.edu/owl/general_writing/academic_writing/using_appropriate_language/stereotypes_and_biased_language.html

Choosing your RC Texts

Which texts not to choose (*The Manual for Language Test Development and Examining, Language,* Policy Division Strasbourg, October 2002):

> Reasons for rejecting texts could include:
>
> - too great an assumption of cultural or local knowledge (unless this is being specifically tested);
> - unsuitable topics, such as war, death, politics and religious beliefs, which may offend or distress some candidates;
> - topics outside the experience of candidates' likely age-group;
> - too high a level of difficulty of vocabulary or concept;
> - technical or stylistic faults or idiosyncrasies;
> - poor editing of the original text. (p. 17)

Developing Essay Topics

- Decide on the essay genre. Do you want your students to report information, analyze the text, argue a position, tell a story, explain a process, or write a proposal, reflection, or evaluation? Even if you want them to write in a mixed genre, their essays should have a clear structure.
- What are some of the real-life situations that will require your students to write in those genres?
- Avoid topics that require your students to write one-track social, political, or religious lectures. Do not push your own point of view in the topic articulation.
- Do not come up with the topics that will urge your students to write insincere essays just to get your approval and a good grade.

List of Tables

Titel der Hausarbeit

Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text

Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text

Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text

Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text

Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text

Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text

Max Mustermann (Beispiel)
max.mustermann@tu-dortmund.de

Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text Text Text Text Text Text Text Text

Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text Text Text Text Text Text Text Text

Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text Text Text Text Text Text Text Text

Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text Text Text Text Text Text Text Text

Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text Text Text Text Text Text Text Text

Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text Text Text Text Text Text Text Text

Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text
Text Text

Max Mustermann (Beispiel)
max.mustermann@tu-dortmund.de

Exemplarisch für B2 Niveau

Schriftart:
Akkurat Light, Akkurat, Arial, Times New Roman (wählen Sie selbst)

Schriftgröße:
12 Punkt

Zeilenabstand:
Einzeilig *

Seitenrand:
Word Office Einstellung: Normal
Rand links: 2,5 cm
Rand rechts: 2,5 cm

Seiten
2 – 2 ½

Wörter
Ca. 1000 – 1200 Wörter

Exemplarisch für C1 Niveau

Schriftart:
Akkurat Light, Akkurat, Arial, Times New Roman (wählen Sie selbst)

Schriftgröße:
12 Punkt

Zeilenabstand:
Einfach *

Seitenrand:
Word Office Einstellung: Normal
Rand links: 2,5 cm
Rand rechts: 2,5 cm

Seiten
3 – 3 ½

Wörter
Ca. 1500 – 1800 Wörter

Anmerkungen:

- Bei der Wahl eines Zeilenabstands von 1,5 Zeilen erhöht sich die Seitenanzahl entsprechend.
- Die Vorlage kann vom Layout geändert werden – eine Fußzeile und/oder Kopfzeile mit den hier angeführten Angaben sollte vorhanden sein.

Auf Wiedersehen Goodbye Au revoir Adiós Arrivederci

zhb Bereich Fremdsprachen - Wegweiser

Postanschrift:
Zentrum für HochschulBildung
Bereich Fremdsprachen
TU Dortmund
Emil-Figge Str. 61, 2. OG
44227 Dortmund

Webseite:
https://fs.zhb.tu-dortmund.de/

Ansprechpersonen

Lehrgebiet Englisch

Dr. Meni Syrou
Leitung zhb Bereich Fremdsprachen
E-Mail: meni.syrou@tu-dortmund.de

Dr. Kai Herklotz
Lehrgebiet Englisch
Lehrgebiet Deutsch als Fremdsprache
E-Mail: kai.herklotz@tu-dortmund.de

Dr. Tetyana Müller-Lyaskovets
Lehrgebiet Englisch
Schreibwerkstatt w.space
E-Mail: tetyana.lyaskovets@tu-dortmund.de

Eun Young Park, M.Ed.
Lehrgebiet Englisch (in Elternzeit)
E-Mail: eunyoung.park@tu-dortmund.de

IMPRESSUM:
Dr. Meni Syrou

Grafiken:
ITMC TU Dortmund und
Kalaudia

Veröffentlichung: März 2023