

## **AG 1: Erstellen eines Rasters zur Bewertung der Klausur schriftliche Textproduktion (sprachübergreifend)**

Marika Fleischer (Universität Rostock) & Johann Fischer (Universität Göttingen)

Ziel der AG war es, eine Art universelles Raster zu erstellen, anhand dessen die Prüfungsleistung schriftliche Textproduktion auf den verschiedenen Niveaustufen und unabhängig von der konkreten Aufgabenstellung bewertet werden kann. Die Entwicklung eines solchen Rasters soll mittelfristig dazu beitragen, die Bewertung der Prüfungen in den verschiedenen Institutionen an gleichen Standards zu orientieren und dadurch zu einer stärkeren Vereinheitlichung der Prüfungsbewertung beitragen.

Als Diskussionsgrundlagen dienten unsere eigenen Erfahrungen und Traditionen, Bewertungsraster, die von einer Reihe von Institutionen zur Verfügung gestellt wurden, die Vorgaben des GER, das dazu gehörige Manual sowie Bewertungsrichtlinien und -raster international bekannter Tests.

Zuerst beleuchteten wir die in den Einrichtungen genutzten Testkriterien<sup>1</sup>. Im Mittelpunkt standen dabei insbesondere ihre Funktion und Wichtung sowie der Vergleich mit dem im GER gebotenen Material<sup>2</sup> und den im Manual zur Verfügung gestellten Anleitungen, Hilfsmitteln, Rastern und Formularen.

Die Diskussion in den Gruppen wurde von folgenden wesentlichen Fragestellungen geprägt:

- Welche Kriterien sind bei einer bestimmten Aufgabenstellung (vor dem Hintergrund eines bestimmten Lebensbereiches) für die Bewertung der schriftlichen Ausdrucksfähigkeit tatsächlich relevant?
- Welche Kompetenz muss für ein bestimmtes Kriterium auf einer bestimmten Niveaustufe erreicht werden?

Als Ergebnis der abschließenden Diskussion schlägt die AG die folgenden drei Kriterien als Basis für die Bewertung der schriftlichen Ausdrucksfähigkeit vor:

### **Inhaltliche Umsetzung** **Pragmatische Kompetenz** **Sprachliche Kompetenz**

---

<sup>1</sup> Auswahl/willkürlich geordnet: globale Wertung, Inhalt, inhaltliche Vollständigkeit, inhaltliche Richtigkeit, Relevanz, Textaufbau / Organisation, Themenentwicklung, Gliederung, Kohärenz und Kohäsion, sprachliche Korrektheit, Wortschatzspektrum, Ausdruck / Stil, Register, Interpunktion / Orthographie, Aufgabenbewältigung, Beschreibung, Argumentation, Nutzung Quellmaterialien

<sup>2</sup> GER

- Globalskalen (S. 35)
- Raster Selbstbeurteilung (S. 36)
- Deskriptoren (S. 66-68, 85-87, (98))
- Kompetenzen (S. 103ff.):
  - linguistische Kompetenzen (S. 110)
  - lexikalische Kompetenzen (S. 111ff)
  - grammatische Kompetenzen (S. 113f)
  - orthographische Kompetenzen (S. 117ff)
  - soziolinguistische Kompetenzen (S. 118ff)
  - pragmatische Kompetenzen (S. 123ff)
  - funktionale Kompetenzen (S. 125ff)

Die folgende Übersicht zeigt, welche Aspekte die einzelnen Kriterien beinhalten und wie sie zueinander gewichtet werden können.

- Einigkeit bestand darin, dass die sprachliche Kompetenz einen höheren Stellenwert erhalten soll und die inhaltliche Umsetzung sowie die pragmatische Kompetenz gleichgewichtig sein sollen.
- Alle Teilnehmer waren sich auch darin einig, dass die sprachliche Korrektheit als ein Aspekt der sprachlichen Kompetenz umso besser/ höher sein muss, je höher die Niveaustufe ist. Mit anderen Worten: Auf der Stufe UNICert III dürften generell weniger sprachliche Fehler gemacht werden als auf der Stufe II. Das führte zu dem Vorschlag, die sprachliche Kompetenz auf der UNICert III-Stufe höher zu wichten als auf der Stufe II. Dieser Ansatz wurde jedoch nicht von allen geteilt.
- Unterschiedliche Auffassungen gab es außerdem zur Art der Bewertung der sprachlichen Korrektheit. Während einige die Bewertung mit Hilfe von Fehlerzahl und Punkten favorisierten, plädierte die Mehrheit für eine verbale Beurteilung in Anlehnung an die Notenskala.

|  |   |
|--|---|
| <p><b>Inhalt</b><br/>30%/25%</p> <ul style="list-style-type: none"> <li>• Vollständigkeit</li> <li>• Korrektheit</li> <li>• Relevanz</li> <li>• Ggf. Originalität, Substanz</li> </ul> <p><b>Pragmatische Kompetenz</b><br/>30%/25%</p> <ul style="list-style-type: none"> <li>• Funktionalität: Wirkung auf Adressaten (inkl. Textsortenadäquatheit)</li> <li>• Textorganisation / Kohärenz / Kohäsion</li> </ul> | <p><b>Sprachliche Kompetenz</b><br/>40%/50%</p> <ul style="list-style-type: none"> <li>• Sprachliche Korrektheit <ul style="list-style-type: none"> <li>➤ Grammatik</li> <li>➤ Wortschatz</li> <li>➤ Orthographie / Interpunktion)</li> </ul> </li> <li>• Variabilität der sprachlichen Mittel <ul style="list-style-type: none"> <li>➤ Grammatik</li> <li>➤ Wortschatz</li> </ul> </li> <li>• Stil / Register</li> </ul> |
|--|---|

Mit diesem Vorschlag ist ein Grundstein für die weitere Arbeit an der Verfeinerung des Bewertungsrasters gelegt.

Dabei werden folgende Aspekte besonders berücksichtigt werden müssen:

- Wichtung der einzelnen Kriterien zueinander
- Auswahl/ Entscheidung für ein Bewertungsprinzip (z.B. Notenskala oder Punkteskala) einschließlich der Festlegung der Grenze zu „nicht bestanden“
- Binnendifferenzierung innerhalb der einzelnen Kriterien, wobei ein Bezug sowohl zu den Kann-Beschreibungen des GER auf der jeweiligen Niveaustufe als auch zum Bewertungsprinzip (Noten oder Punkte) hergestellt werden muss
- Differenzierung zwischen den einzelnen Niveaustufen (z.B. UNICert II und UNICert III)

**Weiteres Vorgehen:**

Die AG-Leiter werden die erarbeiteten Vorschläge und Vorlagen zusammenführen und zu einem ersten Vorschlag ausarbeiten, der anschließend an alle Teilnehmer zur Kommentierung und Überarbeitung in der Gruppe (z.B. über ein Wiki) verschickt wird.

**Nachtrag:**

Inzwischen arbeitet eine Arbeitsgruppe der Universität Göttingen an einem Entwurf für die verschiedenen Stufen basierend auf den Vorarbeiten aus Rostock. Hierzu hat in einem ersten Schritt eine wiss. Hilfskraft die Vorlagen zusammengeführt, die Kann-Beschreibungen aus dem GER etc. integriert und zu einem Bewertungsraster für die verschiedenen Stufen ausgearbeitet. Diese Vorlage wurde dann von Lehrkräften aus verschiedenen Sprachbereichen überarbeitet. Sobald ein einsatzfähiger Entwurf vorliegt, soll er an die AG-Mitglieder zur Fertigstellung weitergeleitet werden.

# Prüfungsprotokoll (A1, A2, B1)

|                     |                 |
|---------------------|-----------------|
| Sprache/Kurs: _____ | Name: _____     |
| Semester: _____     | Vorname: _____  |
| Dozent/in: _____    | Matr.Nr.: _____ |

**Schriftliche Prüfung:**

**Datum:** \_\_\_\_\_

| Inhaltliche Gestaltung der Texte   | 1 | 2 | 3 | 4 | 5 |
|--|---|---|---|---|---|
| <b>Aufgabenbewältigung (Vollständigkeit des Textes):</b> <ul style="list-style-type: none"> <li>- Der/Die Studierende erkannte die Aufgabenstellung und konnte diese angemessen bewältigen.</li> <li>- Der/Die Studierende kann Meinungen, Pläne und Handlungen kurz schriftlich begründen und erklären.</li> <li>- Der/Die Studierende kann in persönlichen Briefen Erfahrungen, Gefühle und Ereignisse detailliert beschreiben.</li> </ul> |   |   |   |   |   |
| <b>Korrektheit:</b> <ul style="list-style-type: none"> <li>- Die inhaltliche Gestaltung entsprach der Aufgabe.</li> <li>- Die Inhalte waren sachlich richtig dargestellt.</li> </ul>   |   |   |   |   |   |
| <b>Relevanz:</b> <ul style="list-style-type: none"> <li>- Die Inhalte waren relevant und aussagekräftig.</li> </ul>  |   |   |   |   |   |
| <b>Originalität:</b> <ul style="list-style-type: none"> <li>- Der/Die Studierende ist von Standardsätzen abgewichen.</li> </ul>  |   |   |   |   |   |
| <b>Bewertung der inhaltlichen Gestaltung des Textes:</b>   |   |   |   |   |   |

| Pragmatische Kompetenz   | 1 | 2 | 3 | 4 | 5 |
|--|---|---|---|---|---|
| <b>Funktionalität: Wirkung auf Adressaten (inkl. Textsortenadäquatheit):</b> <ul style="list-style-type: none"> <li>- Der/Die Studierende kann kurze Berichte schreiben, in welchen Alltagsinformationen und Gründe für Handlungsweisen angegeben werden.</li> <li>- Der/Die Studierende kann die wichtigsten Einzelheiten eines unvorhersehbaren Ereignisses, wie z.B. eines Unfalls, beschreiben.</li> <li>- Der/Die Studierende kann Hoffnungen und Wünsche äußern.</li> <li>- Der/Die Studierende kann, z.B. bei Problemen, Anfragen entgegennehmen und dazu Notizen anfertigen.</li> <li>- Der/Die Studierende kann die Handlung eines Buches oder Films wiedergeben und die eigene Reaktion darauf beschreiben.</li> <li>- Der/Die Studierende kann Standardbriefe schreiben, um detaillierte Informationen zu ver- bzw. ermitteln (z.B. auf eine Anzeige antworten).</li> </ul> |   |   |   |   |   |
| <b>Textorganisation / Kohärenz / Kohäsion:</b> <ul style="list-style-type: none"> <li>- Der Text war klar strukturiert.</li> <li>- Die Einleitung war gelungen.</li> <li>- Der/Die Studierende kann über Themen, die ihm/ihr vertraut sind oder ihn/sie persönlich interessieren, verständliche und zusammenhängende Texte schreiben.</li> </ul>   |   |   |   |   |   |
| <b>Bewertung der pragmatischen Kompetenz:</b>  |   |   |   |   |   |

| Sprachliche Kompetenz  | 1 | 2 | 3 | 4 | 5 |
|--|---|---|---|---|---|
| <b>Sprachliche Korrektheit:</b>  |   |   |   |   |   |
| <b>Grammatik:</b><br>- Die Syntax war korrekt.<br>- Kontextbezogene Konstruktionen wurden richtig verwendet.             |   |   |   |   |   |
| <b>Wortschatz:</b><br>- Der/Die Studierende überzeugte durch den Einsatz eines reichen Wortschatzes.                     |   |   |   |   |   |
| <b>Orthographie / Interpunktion:</b><br>- Rechtschreibung und Zeichensetzung ist korrekt.                                |   |   |   |   |   |
| <b>Variabilität der sprachlichen Mittel:</b>   |   |   |   |   |   |
| <b>Grammatik:</b><br>- Der/Die Studierende setzt unterschiedliche Satzkonstruktionen korrekt ein.                        |   |   |   |   |   |
| <b>Wortschatz:</b><br>- Der/Die Studierende kann komplizierte Sachverhalte zu vertrauten Themen beschreiben.             |   |   |   |   |   |
| <b>Register / Stil:</b><br>- Der/Die Studierende kann zu vertrauten allgemeinen Themen stilistisch angemessen schreiben. |   |   |   |   |   |
| <b>Bewertung der sprachlichen Kompetenz:</b>   |   |   |   |   |   |

|   | Berechnung   | Note |
|---|--------------|------|
| Inhaltliche Gestaltung der Texte (25 %)       | _____ x 25 % |      |
| Pragmatische Kompetenz (25 %)                 | _____ x 25 % |      |
| Sprachliche Kompetenz (50 %)                  | _____ x 50 % |      |
| <b>Endnote Schriftlicher Ausdruck (100 %)</b> |              |      |

**Unterschrift:** \_\_\_\_\_

| Noten                  | Verbale Charakterisierung  |
|------------------------|--|
| 1,0 – 1,3              | Schreibt korrekt gemäß der grammatischen, lexikalischen und orthographischen Normen der Zielsprache.   |
| 1,7 – 2,0 – 2,3        | Schreibt korrekt mit ganz geringfügigen Fehlern.   |
| 2,7 – 3,0 – 3,3        | Schreibt weitgehend korrekt mit geringfügigen Fehlern.   |
| 3,7 - 4,0              | Schreibt teilweise mit geringfügigen Fehlern, aber auch mit einzelnen größeren Fehlern, genügt jedoch insgesamt noch den gestellten Anforderungen.               |
| 5,0<br>Nicht bestanden | Schreibt mit zahlreichen sinnentstellenden Fehlern und zahlreichen Verstößen gegen die grammatischen, lexikalischen und orthographischen Normen der Zielsprache. |

# Prüfungsprotokoll (B2.1, B2.2)

|                     |                 |
|---------------------|-----------------|
| Sprache/Kurs: _____ | Name: _____     |
| Semester: _____     | Vorname: _____  |
| Dozent/in: _____    | Matr.Nr.: _____ |

**Schriftliche Prüfung:**

**Datum:** \_\_\_\_\_

| Inhaltliche Gestaltung der Texte   | 1 | 2 | 3 | 4 | 5 |
|--|---|---|---|---|---|
| <b>Aufgabenbewältigung (Vollständigkeit des Textes):</b> <ul style="list-style-type: none"> <li>- Der/Die Studierende erkannte die Aufgabenstellung und konnte diese angemessen bewältigen.</li> <li>- Der/Die Studierende bewies die Fähigkeit, die Fragestellung sachlich richtig zu erarbeiten.</li> <li>- Die Inhalte wurden klar und überzeugend dargestellt.</li> <li>- Der/Die Studierende kann über eine Vielzahl von vertrauten Themen klare und detaillierte Texte schreiben.</li> </ul> |   |   |   |   |   |
| <b>Korrektheit der Inhalte:</b> <ul style="list-style-type: none"> <li>- Die inhaltliche Gestaltung entsprach der Aufgabe.</li> <li>- Die Inhalte waren sachlich richtig dargestellt.</li> <li>- Einzelne Aspekte wurden ggf. kritisch beleuchtet.</li> <li>- Der/Die Studierende kann Ursachen, Folgen und hypothetische Situationen kritisch analysieren.</li> </ul>   |   |   |   |   |   |
| <b>Relevanz der Inhalte:</b> <ul style="list-style-type: none"> <li>- Die Inhalte waren relevant und aussagekräftig und wurden mit geeigneten Beispielen verdeutlicht.</li> </ul>  |   |   |   |   |   |
| <b>Originalität:</b> <ul style="list-style-type: none"> <li>- Der/Die Studierende konnte durch eine originelle Gestaltung der Inhalte die Aufmerksamkeit des Lesers gewinnen.</li> </ul>   |   |   |   |   |   |
| <b>Bewertung der inhaltlichen Gestaltung des Textes:</b>   |   |   |   |   |   |

| Pragmatische Kompetenz   | 1 | 2 | 3 | 4 | 5 |
|--|---|---|---|---|---|
| <b>Funktionalität: Wirkung auf Adressaten (inkl. Textsortenadäquatheit):</b> <ul style="list-style-type: none"> <li>- Der/Die Studierende kann unterschiedliche Textsorten schreiben.</li> <li>- Der/Die Studierende kann offizielle Briefe, E-Mails und Berichte schreiben.</li> <li>- Der/Die Studierende kann Briefe schreiben, in denen er/sie seine/ihre Gefühle differenziert beschreibt, die persönliche Bedeutung von Ereignissen und Erfahrungen hervorhebt und die Nachrichten und Standpunkte seines/ihrer Briefpartners kommentiert.</li> <li>- Der/Die Studierende kann Neuigkeiten, Standpunkte und Gefühle schriftlich darstellen und zu denen anderer Stellung nehmen.</li> <li>- Der/Die Studierende kann in einem Aufsatz oder Bericht etwas erörtern, dabei Gründe für oder gegen einen bestimmten Standpunkt angeben und die Vor- und Nachteile verschiedener Optionen abwägen.</li> <li>- Der/Die Studierende kann eine kurze Rezension eines Films oder Buches schreiben.</li> </ul>   |   |   |   |   |   |
| <b>Textorganisation / Kohärenz / Kohäsion / Layout:</b> <ul style="list-style-type: none"> <li>- Der Text war klar strukturiert.</li> <li>- Die Einleitung war gelungen.</li> <li>- Die Strukturierung, Verknüpfung der Argumente und Übergänge waren gelungen.</li> <li>- Inhaltliche Schwerpunkte wurden klar formuliert.</li> <li>- Der/Die Studierende kann unterschiedliche Ideen und Lösungen für ein Problem gegeneinander abwägen.</li> <li>- Der/Die Studierende kann Informationen und Argumente aus unterschiedlichen Quellen zusammenfassen.</li> <li>- Der/Die Studierende kann in leicht lesbarer Form klare, detaillierte, zusammenhängende Beschreibungen realer oder fiktiver Ereignisse und Erfahrungen verfassen und dabei den Zusammenhang zwischen verschiedenen Ideen deutlich machen.</li> <li>- Der Schluss war gelungen.</li> <li>- Der/Die Studierende kann in einem Aufsatz oder Bericht Informationen wiedergeben oder Argumente für oder gegen einen bestimmten Standpunkt darlegen.</li> <li>- Layout und Absätze sind korrekt und unterstützen die Textorganisation.</li> </ul> |   |   |   |   |   |
| <b>Bewertung der pragmatischen Kompetenz:</b>  |   |   |   |   |   |

| Sprachliche Kompetenz   | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| <b>Sprachliche Korrektheit:</b>   |   |   |   |   |   |
| <b>Grammatik:</b><br>- Die Syntax war korrekt.<br>- Kontextbezogene Konstruktionen wurden korrekt verwendet.  |   |   |   |   |   |
| <b>Wortschatz:</b><br>- Der/Die Studierende überzeugte durch den Einsatz eines reichen Wortschatzes.<br>- Der/Die Studierende kann Verbindungswörter sinnvoll verwenden, um logische Beziehungen zu verdeutlichen.  |   |   |   |   |   |
| <b>Orthographie / Interpunktion:</b><br>- Rechtschreibung und Zeichensetzung sind korrekt.  |   |   |   |   |   |
| <b>Variabilität der sprachlichen Mittel:</b>  |   |   |   |   |   |
| <b>Grammatik:</b><br>- Der/Die Studierende überzeugt durch eine umfassende grammatische Flexibilität.   |   |   |   |   |   |
| <b>Wortschatz:</b><br>- Der/Die Studierende kann Standpunkte, Gefühle und Erfahrungen nuanciert auszudrücken.<br>- Der/Die Studierende kann Neuigkeiten und Standpunkte erfolgreich vermitteln.<br>- Der/Die Studierende kann unterschiedliche zusammenhängende Texte zu Themen aus seinem/ihrer persönlichen, studien- oder berufsbezogenen Interessensgebiet verfassen. |   |   |   |   |   |
| <b>Register / Stil:</b><br>- Der/Die Studierende kann Texte zu persönlichen, studien- oder berufsbezogenen Themen sowie Sachinformationen aus verschiedenen Quellen und Medien zusammenfassen.<br>- Der/Die Studierende kann standardisierte offizielle Briefe schreiben, unter Verwendung des angemessenen Registers und unter Beachtung der Gepflogenheiten.            |   |   |   |   |   |
| <b>Bewertung der sprachlichen Kompetenz:</b>  |   |   |   |   |   |

|   | Berechnung  | Note |
|---|-------------|------|
| Inhaltliche Gestaltung der Texte (25 %)       | _____ x 25% |      |
| Pragmatische Kompetenz (25 %)                 | _____ x 25% |      |
| Sprachliche Kompetenz (50 %)                  | _____ x 50% |      |
| <b>Endnote Schriftlicher Ausdruck (100 %)</b> |             |      |

**Unterschrift:** \_\_\_\_\_

| Noten                  | Verbale Charakterisierung  |
|------------------------|--|
| 1,0 – 1,3              | Schreibt korrekt gemäß der grammatischen, lexikalischen und orthographischen Normen der Zielsprache.   |
| 1,7 – 2,0 – 2,3        | Schreibt korrekt mit ganz geringfügigen Fehlern.   |
| 2,7 – 3,0 – 3,3        | Schreibt weitgehend korrekt mit geringfügigen Fehlern.   |
| 3,7 - 4,0              | Schreibt teilweise mit geringfügigen Fehlern, aber auch mit einzelnen größeren Fehlern, genügt jedoch insgesamt noch den gestellten Anforderungen.               |
| 5,0<br>Nicht bestanden | Schreibt mit zahlreichen sinnentstellenden Fehlern und zahlreichen Verstößen gegen die grammatischen, lexikalischen und orthographischen Normen der Zielsprache. |

# Prüfungsprotokoll (C1.1, C1.2)

|                     |                 |
|---------------------|-----------------|
| Sprache/Kurs: _____ | Name: _____     |
| Semester: _____     | Vorname: _____  |
| Dozent/in: _____    | Matr.Nr.: _____ |

## Schriftliche Prüfung:

Datum: \_\_\_\_\_

| Inhaltliche Gestaltung der Texte   | 1 | 2 | 3 | 4 | 5 |
|--|---|---|---|---|---|
| <b>Aufgabenbewältigung (Vollständigkeit des Textes):</b> <ul style="list-style-type: none"> <li>- Der/Die Studierende erkannte die Aufgabenstellung und konnte diese angemessen bewältigen.</li> <li>- Der/Die Studierende bewies die Fähigkeit, die Fragestellung sachlich richtig zu erarbeiten.</li> <li>- Die Inhalte wurden klar und überzeugend dargestellt.</li> <li>- Die eigene Meinung bzw. Perspektive wurde nuanciert und überzeugend dargestellt.</li> <li>- Der/Die Studierende konnte generelle Kenntnisse auf das gewählte Thema anwenden.</li> <li>- Der/Die Studierende kann komplexe Sachverhalte klar und detailliert schriftlich darstellen.</li> </ul> |   |   |   |   |   |
| <b>Korrektheit der Inhalte:</b> <ul style="list-style-type: none"> <li>- Die inhaltliche Gestaltung entsprach der Aufgabe.</li> <li>- Die Inhalte waren sachlich richtig dargestellt.</li> <li>- Einzelne Aspekte wurden ggf. kritisch beleuchtet bzw. analysiert.</li> <li>- Der vorgestellte Lösungsansatz war plausibel und überzeugend.</li> <li>- Es wurden gute Beispiele angeführt.</li> <li>- Der/Die Studierende kann in komplexen formellen Briefen seine/ihre Argumente klar und wirkungsvoll darlegen (z.B. eine Beschwerde vorbringen, eine kontroverse Stellung zu einer Problemfrage beziehen).</li> </ul>  |   |   |   |   |   |
| <b>Relevanz der Inhalte:</b> <ul style="list-style-type: none"> <li>- Die Inhalte waren relevant und aussagekräftig.</li> <li>- Der/Die Studierende kann in Briefen, Aufsätzen oder Berichten über komplexe Sachverhalte schreiben und die für ihn/sie wesentlichen Aspekte hervorheben.</li> <li>- Der/Die Studierende kann Standpunkte durch zusätzliche Argumente, Begründungen und sinnvolle Beispiele weiter ausführen und untermauern.</li> </ul>  |   |   |   |   |   |
| <b>Originalität / Substanz:</b> <ul style="list-style-type: none"> <li>- Der/Die Studierende konnte durch eine originelle Gestaltung der Inhalte die Aufmerksamkeit des Lesers gewinnen.</li> <li>- Der/Die Studierende hat zahlreiche relevante Informationen in seiner Darstellung verarbeitet.</li> </ul>   |   |   |   |   |   |
| <b>Bewertung der inhaltlichen Gestaltung des Textes:</b>   |   |   |   |   |   |

| Pragmatische Kompetenz  | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| <b>Funktionalität: Wirkung auf Adressaten (inkl. Textsortenadäquatheit):</b> <ul style="list-style-type: none"> <li>- Der/Die Studierende kann unterschiedliche Textsorten schreiben.</li> <li>- Der/Die Studierende kann offizielle Briefe, E-Mails und Berichte schreiben.</li> <li>- Der/Die Studierende kann Protokolle erstellen.</li> <li>- Der/Die Studierende kann Standardbriefe schreiben, um detaillierte Informationen zu ver- bzw. ermitteln (z.B. auf eine Anzeige antworten, sich auf eine Stelle bewerben).</li> <li>- Der/Die Studierende kann klare, gut strukturierte und ausführliche Beschreibungen oder eigene fiktionale Texte in lesergerechtem, überzeugendem, persönlichem und natürlichem Stil verfassen.</li> </ul> |   |   |   |   |   |
| <b>Textorganisation / Kohärenz / Kohäsion / Layout:</b> <ul style="list-style-type: none"> <li>- Der Text war klar strukturiert.</li> <li>- Die Einleitung war gelungen.</li> <li>- Die Strukturierung, Verknüpfung der Argumente und Übergänge waren gelungen.</li> <li>- Inhaltliche Schwerpunkte wurden klar formuliert und unterstützende Einzelheiten angeführt.</li> <li>- Der/Die Studierende kann klar und flüssig schreiben.</li> <li>- Der/Die Studierende kann die eigene Ansicht ausführlich darstellen.</li> <li>- Der Schluss war gelungen.</li> <li>- Layout und Absätze sind korrekt und unterstützen die Textorganisation.</li> </ul>  |   |   |   |   |   |
| <b>Bewertung der pragmatischen Kompetenz:</b>   |   |   |   |   |   |



| Sprachliche Kompetenz   | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| <b>Sprachliche Korrektheit:</b>   |   |   |   |   |   |
| <b>Grammatik:</b> <ul style="list-style-type: none"> <li>- Die Syntax war korrekt.</li> <li>- Kontextbezogene Konstruktionen wurden richtig verwendet.</li> </ul>   |   |   |   |   |   |
| <b>Wortschatz:</b> <ul style="list-style-type: none"> <li>- Der/Die Studierende überzeugte durch den Einsatz eines reichen Wortschatzes.</li> <li>- Fachsprachliche Terminologie wurde korrekt verwendet.</li> <li>- Der/Die Studierende kann zahlreiche Verbindungswörter sinnvoll verwenden, um logische Beziehungen zu verdeutlichen.</li> </ul>   |   |   |   |   |   |
| <b>Orthographie / Interpunktion:</b> <ul style="list-style-type: none"> <li>- Die Rechtschreibung ist korrekt.</li> <li>- Die Zeichensetzung ist korrekt.</li> </ul>  |   |   |   |   |   |
| <b>Variabilität der sprachlichen Mittel:</b>  |   |   |   |   |   |
| <b>Grammatik:</b> <ul style="list-style-type: none"> <li>- Der/Die Studierende beweist große grammatikalische Flexibilität.</li> <li>- Der/Die Studierende verwendet komplexe Strukturen.</li> </ul>  |   |   |   |   |   |
| <b>Wortschatz:</b> <ul style="list-style-type: none"> <li>- Der/Die Studierende überzeugte durch den Einsatz eines reichen allgemein- und fachsprachlichen Wortschatzes.</li> <li>- Der/Die Studierende kann sich präzise und differenziert ausdrücken,</li> <li>- Der/Die Studierende verwendet Synonyme angemessen.</li> <li>- Der/Die Studierende kann graduelle Unterschiede von Meinungen und Aussagen z.B. in Bezug auf Sicherheit/Unsicherheit, Überzeugung/Zweifel, Wahrscheinlichkeit formulieren.</li> <li>- Der/Die Studierende kann Standpunkte, Gefühle und Erfahrungen nuanciert ausdrücken.</li> </ul> |   |   |   |   |   |
| <b>Register / Stil:</b> <ul style="list-style-type: none"> <li>- Der/Die Studierende kann unter Verwendung des angemessenen Registers den Stil wählen, der für die jeweiligen Leser adäquat ist.</li> <li>- Der/Die Studierende kann sein/ihr Vokabular und den eigenen Stil dem Kontext anpassen.</li> <li>- Der/Die Studierende verwendet das Register beruflicher bzw. universitärer sprachlicher Anforderungen angemessen.</li> </ul>   |   |   |   |   |   |
| <b>Bewertung der sprachlichen Kompetenz:</b>  |   |   |   |   |   |

|   | Berechnung  | Note |
|---|-------------|------|
| Inhaltliche Gestaltung der Texte (25 %)       | _____ x 25% |      |
| Pragmatische Kompetenz (25 %)                 | _____ x 25% |      |
| Sprachliche Kompetenz (50 %)                  | _____ x 50% |      |
| <b>Endnote Schriftlicher Ausdruck (100 %)</b> |             |      |

**Unterschrift der Prüfer:** \_\_\_\_\_

| Noten                  | Verbale Charakterisierung  |
|------------------------|--|
| 1,0 – 1,3              | Schreibt korrekt und stilistisch angemessen, gemäß den grammatischen, lexikalischen und orthographischen Normen der Zielsprache zu allgemeinen und / oder fachspezifischen Themen. |
| 1,7 – 2,0 – 2,3        | Schreibt korrekt und stilistisch angemessen, jedoch mit vereinzelten Fehlern, zu allgemeinen und / oder fachspezifischen Themen.   |
| 2,7 – 3,0 – 3,3        | Schreibt verständliche Texte, jedoch mit vereinzelten Fehlern in Wortschatz und Grammatik zu allgemeinen und / oder fachspezifischen Themen.                                       |
| 3,7 - 4,0              | Schreibt Texte mit beschränktem Wortschatz und gröberen Fehlern in der Grammatik, genügt aber insgesamt den Anforderungen.   |
| 5,0<br>Nicht bestanden | Schreibt Texte mit beschränktem Wortschatz, zahlreichen sinnentstellenden Fehlern und gröberen Verstößen gegen die grammatischen Normen der Zielsprache.                           |

# **Bewertung der Klausur schriftliche Textproduktion anhand von Prüfungsbeispielen (Englisch, UNiCert® II und III)**

## **AG-2.1**

**John Nixon, Angelika Grigor**

Die Arbeitsgruppe hat sich mit drei Hauptthemen beschäftigt: der Definition einer offenen Prüfungsaufgabe im Rahmen des universitären Fremdsprachenunterrichts, mit Kriterien zur Erstellung und Bewertung offener Prüfungsaufgaben, sowie der Bewertung von Prüfungsbeispielen unter Berücksichtigung der von der AG entwickelten Bewertungskriterien.

### **Was sind offene Prüfungsaufgaben?**

Offene Prüfungsaufgaben ermöglichen es den Studierenden, einen freien und ausführlichen Text zu produzieren. Im Gegensatz zu geschlossenen Prüfungsaufgaben hat eine offene Variante nicht eine einzige richtige Lösung. Daher gibt es viele Schwierigkeiten bei der Bewertung offener Prüfungsaufgaben, die anhand von Evaluierungskriterien teilweise behoben werden können.

Zu offenen Prüfungsaufgaben zählen: Freitextaufgaben, Essay-Aufgaben und Briefe. Lückentexte und Multiple-Choice-Fragen sind hingegen geschlossene Aufgabenformen. Im Kontext des universitären Fremdsprachenunterrichts und vor allem in den Fachsprachen werden Berichte, Fallstudien und Beschreibungen von Prozessen als weitere Beispiele von offenen Aufgaben gesehen.

Die von Bloom entworfene Taxonomie des kognitiven Lernens (1956), die eine sechsstufige Skala der Komplexität von Fertigkeiten enthält, wurde im Rahmen des Definierens einer offenen Prüfungsaufgabe genauer geprüft. Die drei komplexesten Fertigkeiten - Evaluation, Synthese, Analyse - wurden von der Mehrheit der AG-Teilnehmer/innen als weitere Merkmale einer offenen schriftlichen Aufgabenform angesehen. Die unteren drei Stufen - Anwendung, Verstehen und Wissen - allein reichen hingegen als Merkmale einer offenen Aufgabe nicht aus, obwohl sie alle bei der Erledigung einer Aufgabe zwangsläufig notwendig sind.

### **Überlegungen bei der Erstellung und Bewertung einer offenen Prüfungsaufgabe**

Sowohl bei der Erstellung einer offenen Prüfungsaufgabe als auch bei deren Bewertung ist es wichtig, einen Katalog von Kriterien zu erstellen, um Qualität, Vergleichbarkeit und Transparenz sicherzustellen.

Ein erstes Kriterium ist selbstverständlich die Frage der „Offenheit“. Anders gesagt, inwieweit ist die erstellte Aufgabe eine offene Variante, ein Thema, mit dem sich die Arbeitsgruppe eine längere Zeit auseinandergesetzt hat.

Neben den sprachlichen Kompetenzen hat sich die Arbeitsgruppe mit weiteren wichtigen Überlegungen bei der Erstellung und Bewertung offener Prüfungsaufgaben befasst, die in der folgenden Tabelle aufgeführt sind:

| Überlegung   | Erläuterung   |
|--|---|
| Hochschulrelevanz  | Ist die Aufgabe typisch und relevant für den Fremdsprachenunterricht im Hochschulbereich? Ist die Antwort des Studierenden für den universitären Bereich angemessen?  |
| Stufenspezifität   | Welcher Stufe gehören sowohl die Aufgabe als auch die Antwort des Studierenden am genauesten an?  |
| Fachsprachenrelevanz   | Ist die Aufgabe allgemeiner Natur oder passt sie zu einer Fachsprache, z.B. <i>Business</i> oder <i>Technical English</i> ? Ist die Aufgabenstellung typisch und authentisch für das spätere Berufsfeld oder für das Studium?   |
| Lernen versus Leistung<br>( <i>achievement</i> versus <i>proficiency</i> ) | Wie bezieht sich die Sprachausbildung auf die Bewertung der Prüfung? Spiegelt sie sich direkt in den Aufgaben wider? Oder dient die Ausbildung eher als eine Übung für die Prüfung, die den Stoff des Unterrichts nicht direkt abfragt? Wenn nicht, inwieweit ist eine offene Prüfungsaufgabe mit dem Abprüfen des im Unterricht gelehrt/gelernten Stoffs zu vereinbaren? |
| Weitere Fertigkeiten /<br>Integrative Fertigkeiten                         | Sind bei der Bewältigung der Aufgabe andere Fertigkeiten, außer rein sprachlichen Kenntnissen, notwendig (z.B. logisches Denken oder die Strukturierung eines technischen Berichts)? Sollten solche weiteren Fertigkeiten überhaupt bewertet werden? Können diese Ansprüche an eine fachsprachen-typische Prüfung gestellt werden?  |
| Inhalt   | Entspricht der Inhalt der Antwort des Studierenden der Stufe und der Aufgabe? Hat der/die Studierende alle Teilfragen bzw. –aufgaben komplett beantwortet?  |

### **Anwendung der erarbeiteten Kriterien auf von UNICert-akkreditierten Einrichtungen gestellte Prüfungen**

Schließlich wurden bereits an UNICert-akkreditierten Einrichtungen erstellte und eingesetzte Prüfungen an die AG-Teilnehmer/innen verteilt und anhand der oben erarbeiteten Kriterien bzw. Überlegungen genauer unter die Lupe genommen. Es hat sich herausgestellt, dass sich offene Prüfungsaufgaben oft schwer mit dem Prüfen des im Unterricht gelernten Stoffs vereinbaren lassen, zumal eine offene Prüfungsaufgabe per Definition dem Studierenden die Entscheidung überlässt, wie er die Aufgabe bewältigt. Zur Problemlösung wurde vorgeschlagen mehrere und verschiedenartige Aufgaben in der schriftlichen Klausur zu

stellen, um den beiden unterschiedlichen Aspekten *achievement* und *performance* gerecht zu werden.

Anschließend wurden die schriftlichen studentischen Antworten zu obigen Aufgaben unter Berücksichtigung der in der AG entwickelten Kriterien analysiert. Zusätzlich wurde das bei den Cambridge-Prüfungen eingesetzte Bewertungsraster, das sieben Kategorien vorsieht - *Accuracy, Content, Organization/Cohesion, Register, Vocabulary, Range* und *Effect* - angewandt, um einen Vergleich zwischen einem etablierten und dem in der AG entwickelten Raster herzustellen.

Obwohl die Sinnhaftigkeit eines Rasters bei der Bewertung der schriftlichen Klausur den meisten AG-Teilnehmer/innen klar war, fand die Idee eines engen Rasters (Punktabzug bei jedem Fehler Punkte) allerdings in der AG aus verschiedenen Gründen keine mehrheitliche Zustimmung. Das Cambridge-Raster ist auf den universitären Kontext nur schwer übertragbar, denn die Fachsprachen und die damit verbundenen, besonderen Prüfungsaufgaben und besonders die Fertigkeiten (z.B. bei einem technischen Bericht) finden in einem für die Allgemeinsprache entwickelten Raster kaum angemessenen Platz; besonders weil in der Bewertung der Allgemeinsprache die Korrektheit der Sprache stärker ins Gewicht fällt als in der Bewertung der berufstypischen Kommunikation.

### **Weitere Überlegungen**

Es wurde für wichtig erachtet das zur Bewertung der Prüfung verwendete Raster den Studierenden im Vorfeld der Prüfung zur Verfügung zu stellen, damit die Erwartungen allen Prüfungskandidaten/innen klar sind. Ebenfalls sollten offene Aufgaben im Curriculum Platz finden, sodass die Prüfung im engeren Verhältnis zur Ausbildung steht. So hätten die Studierenden die Möglichkeit, sich an die für sie oft neuen Prüfungsformen zu gewöhnen.

## AG 2.2 Bewertung der Klausur schriftliche Textproduktion anhand von Prüfungsbeispielen (sprachspezifisch)

### AG 2.2 Französisch (UNlcert® I, II)

Teilnehmende: Andreas Bahr (Frankfurt/O), Frank Hoffmann (Aachen), Christophe Hohwald (Lüneburg), Regina Mügge (Halle-Wittenberg), Cornelia Personne (Darmstadt)

#### Inhalte – einige Diskussionspunkte:

- Die Unterscheidung von Norm- und Kriterienorientierung im Sinne des GER ist begrifflich unglücklich, da Bewertung natürlich immer auf Kriterien basiert. Die Unterschiede zwischen verschiedenen Bewertungs“philosophien“ liegt vielmehr in dem Grad der Explizitheit dieser Kriterien und in der Art der Kriterien selbst. Der Begriff „Normorientierung“ steht dabei für eine Bezugsgruppenorientierung wie sie das traditionelle Notensystem oder das französische „concours“-System repräsentieren, bei denen die Prüflinge in eine Rangfolge gebracht werden, also eine soziale Bezugsnorm ausschlaggebend ist. „Kriterienorientierung“ im engeren Sinn (s.o., vgl. GER) führt dagegen einen externen Bezugsrahmen ein, auf den die Leistungen der Prüflinge „bezogen“ werden. Der externe Bezugsrahmen im Spracherwerbskontext orientiert sich an den Beschreibungen der zu erlangenden Kompetenzen, häufig auf der Grundlage des GER.
- Bewertungskriterien: Es gibt eine Vielzahl von Kriterien, die von professionellen Testanbietern und Sprachlehrforschern zur Bewertung schriftlicher Produktionen vorgeschlagen werden wie auch von den Lehrenden selbst aufgestellte und in der Praxis verwendete Kriterien; häufig finden sich in Kriterienrastern Kriterien wie:
  1. Inhalts(erfüllung) vs. Korrektheit in der sprachlichen Form
    - § Inhalt:
      - § Reaktion auf Stimuli
      - § Erfüllung der inhaltlichen Anforderungen
    - § sprachliche Form:
      - § Kohärenz
      - § Lexik (Reichhaltigkeit und Korrektheit)
      - § morphosyntaktische Strukturen (idem)
      - § Stil
  2. Globaleindruck (globale Bewertung im Gegensatz zu analytischen Kriterien)
  3. Textsortenangemessenheit

- Bewertungsraster: als Diskussionsgrundlage dienten im Workshop folgende Raster:
  1. Raster von DELF B1 und B2
  2. Raster von Telc B1 und B2 (Französisch)
  3. (niveaustufenübergreifendes) Raster des Sprachenzentrums der Universität Rostock (Erprobungsversion)
  4. Bewertungsraster für Französisch UNICert® II der Europa-Universität Viadrina (Frankfurt/O)
  5. (nur als Anschauungsmaterial aus dem Bereich Deutsch: gestufte Raster von TestDaF und vom Deutschen Sprachdiplom DSD)
  
- Die diskutierten Bewertungsraster unterscheiden sich in mehreren Aspekten:
  1. sprachenspezifisch vs. sprachenübergreifend
  2. niveaustufenspezifisch vs. niveaustufenüberschreitend
  3. textsorten- oder gar aufgabenspezifisch
  4. kriterienbezogen (d.h. in der Art der Messung der Kriterien-Erfüllung)
    - § über die Vergabe relativer Punktwerte für spezifische Kriterien (ggf. mit zusätzlicher Gewichtung der Kriterien)
    - § durch Abgleichen mit kriterienspezifischen Deskriptoren für einzelne Bewertungsstufen
  
- Anwendung und Evaluierung der Bewertungsraster zur vergleichenden Bewertung (originaler) studentischer Textproduktionen (Beispiele zu UNICert® I vom ZFA der Ruhr-Universität Bochum, Beispiele zu UNICert® II vom Sprachenzentrum der Universität Rostock) unter dem Blickwinkel folgender Leitfragen zur Reflexion des Bewertungsprozesses:
  1. Entsprechen die in den Kleingruppen auf der Grundlage der verschiedenen Raster erarbeiteten Bewertungen der Klausuren dem subjektiven Eindruck der Workshop-Teilnehmer oder fallen sie besser/ schlechter aus?
  2. Bilden die für die Bewertung auf den Niveaustufen UNICert® I und II genannten Kriterien in den vorgeschlagenen Rastern angemessen das Anforderungsniveau/die Kompetenzen der jeweiligen Niveaustufe ab? Widerspiegeln die Kriterien die Unterschiede zwischen den Niveaustufen?
  3. Welche sonstigen Probleme hinsichtlich der Bewertung haben sich bei der Arbeit mit den Rastern ergeben?
  4. Erfassen die vorgeschlagenen Bewertungsraster sowohl Deskriptoren der Sprachbeherrschung (GER Kapitel 5) als auch Deskriptoren für kommunikative Aktivitäten (GER Kapitel 4)? Ist die Gewichtung dieser Deskriptoren untereinander ausgewogen und der jeweiligen Niveaustufe angemessen?
  5. Sind die Beschreibungen der Notenstufen aussagekräftig und angemessen?

#### Fazit der Diskussion:

Alle Teilnehmer stimmten darin überein, dass holistische und analytische Aspekte gleichermaßen in die Bewertung eingehen sollen, erkannten aber gleichzeitig an, dass die Umsetzung dieser Forderung in den verwendeten Rastern noch nicht optimal gelöst ist (verwendete

Kriterien, Wichtung der Kriterien usw.). Eine Abhilfe können hier – wie insbesondere von professionellen Testanbietern praktiziert – Prüferschulungen, z. B. im Sinne einer (nicht auf eine Zweitkorrektur beschränkten) In-House-Kalibrierung, schaffen, auch wenn es in der Praxis nicht einfach ist, solche Prüferschulungen zu realisieren.

Aufgabenbezogene Bewertungsraster enthalten in der Regel bis ins Detail ausgearbeitete Kriterien, müssen aber an die Spezifika jeder einzelnen Schreibaufgabe angepasst werden. Damit sind der Vergleichbarkeit dieser Raster und der Ergebnisse ihrer Anwendung naturgemäß enge Grenzen gesetzt.

Generell wurde bei der Anwendung der verschiedenen Bewertungsraster deutlich, dass die Beurteilung der Angemessenheit einer Prüfungsleistung nur unter der Voraussetzung möglich ist, dass den Prüfungsteilnehmern der Erwartungshorizont des (angenommenen) Textrezipienten, und damit auch des Prüfers, anhand der Aufgabenstellung hinreichend verdeutlicht wurde. Entsprechend muss die Aufgabenstellung präzise Vorgaben, z. B. zum (hypothetischen) kommunikativen Kontext, enthalten, ohne allerdings dem Prüfungsteilnehmer eine Art „Schreibanleitung“ in die Hand zu geben, die mehr oder weniger „mechanisch“ abzuarbeiten wäre.

#### Zum Nachschlagen:

ALTE (1995 / 2005): Handreichungen für Testautoren. Modul 4: Aspekte der Bewertung und Auswertung (s. [http://www.testdaf.de/aktuelle/pdf/ALTE\\_Deutsche\\_HR\\_Modul4.pdf](http://www.testdaf.de/aktuelle/pdf/ALTE_Deutsche_HR_Modul4.pdf))

Europarat/Rat für kulturelle Zusammenarbeit (2001): Gemeinsamer europäischer Referenzrahmen für Sprachen: lernen, lehren, beurteilen. Berlin etc., Langenscheidt.

#### Anhang:

Bewertungsraster verschiedener standardisierter Prüfungen

Bewertungsraster schriftliche Textproduktion: Sprachenzentrum Universität Rostock

Anhang 1 zum Abschlussbericht AG 2 Französisch:

Bewertungsraster standardisierter Prüfungen (schriftliche Textproduktion)  
(zusammengestellt von Astrid Reich, ZFA, RUB)

I. (auf das Französische bezogen: Bewertungskriterien als Kategorien:)

I.I. DELF (s. <http://www.ciep.fr/de/delfdalf/sujet.php>):

B1:

Themenbeispiel 1 (B1), S. 16 (Druckseite des pdf-Dokuments)

[http://www.ciep.fr/delfdalf/documents/DELF\\_B1.pdf](http://www.ciep.fr/delfdalf/documents/DELF_B1.pdf)

Themenbeispiel 2 (B1), S. 21 (Druckseite des pdf-Dokuments)

[http://www.ciep.fr/delfdalf/documents/DELF\\_B1\\_exemple2.pdf](http://www.ciep.fr/delfdalf/documents/DELF_B1_exemple2.pdf)

B2:

Themenbeispiel (B2), S. 17 (Druckseite des pdf-Dokuments)

[http://www.ciep.fr/delfdalf/documents/DELF\\_B2.pdf](http://www.ciep.fr/delfdalf/documents/DELF_B2.pdf)

I.II Telc (<http://www.telc.net/Francais.404+M54a708de802.0.html>):

B1:

jeu d'épreuves témoin 2:

([http://www.telc.net/Detailansicht.463.0.html?&no\\_cache=1&L=0&tx\\_ttnews\[tt\\_news\]=90&tx\\_ttnews\[backPid\]=129&cHash=7a5a7209a1](http://www.telc.net/Detailansicht.463.0.html?&no_cache=1&L=0&tx_ttnews[tt_news]=90&tx_ttnews[backPid]=129&cHash=7a5a7209a1)) - Druckseiten 40f. des pdf-Dokuments (Achtung:

Seitennummerierung abweichend)

[http://www.telc.net/index.php?eID=tx\\_nawsecuredl&u=0&file=uploads/media/telc\\_Francais\\_B1\\_Modelltest\\_2.pdf&t=1257846261&hash=d0bcef0bbb332898657fbb12ba8657d0](http://www.telc.net/index.php?eID=tx_nawsecuredl&u=0&file=uploads/media/telc_Francais_B1_Modelltest_2.pdf&t=1257846261&hash=d0bcef0bbb332898657fbb12ba8657d0)

B2:

jeu d'épreuves B2:

[http://www.telc.net/Detailansicht.463.0.html?&no\\_cache=1&L=0&tx\\_ttnews\[tt\\_news\]=104&tx\\_ttnews\[backPid\]=129&cHash=ae8b335835](http://www.telc.net/Detailansicht.463.0.html?&no_cache=1&L=0&tx_ttnews[tt_news]=104&tx_ttnews[backPid]=129&cHash=ae8b335835) - Druckseiten 43f. des pdf-Dokuments (Achtung:

Seitennummerierung abweichend)

[http://www.telc.net/index.php?eID=tx\\_nawsecuredl&u=0&file=uploads/media/telc\\_Francais\\_B2\\_Modelltest\\_1\\_01.pdf&t=1257846223&hash=11153afd67cd23a422f70d9a71b0b30f](http://www.telc.net/index.php?eID=tx_nawsecuredl&u=0&file=uploads/media/telc_Francais_B2_Modelltest_1_01.pdf&t=1257846223&hash=11153afd67cd23a422f70d9a71b0b30f)



**Anhang 2: Bewertungsbogen schriftliche Textproduktion (bezogen auf Erwartungshorizont der jeweiligen Niveaustufe) – Uni Rostock/SZ**

| Note              | Pragmatische Kompetenz   | Inhalt  | Sprachliche Kompetenz   |  |
|-------------------|--|---|---|--|
|                   | Themenentwicklung bezogen auf Erwartungshorizont, Textorganisation   |   | Ausdruck (Lexik, Idiomatik, Textverknüpfungs-/Redemittel, syntaktische Angemessenheit)  | Morphologisch-syntaktische Korrektheit   |
| sehr gut          | Text durchgängig flüssig lesbar u. klar strukturiert; Gedankengang logisch u. problemlos nachvollziehbar; gewünschte Wirkung wird voll erzielt                               | Text wird Aufgabenstellung gerecht. Lt. Aufgabenstellung implizit enthaltene od. explizit vorgegebene Punkte werden angemessen ( sachl. richtig, zielführend) behandelt | variabler und präziser Gebrauch, der Textsorte angemessene syntakt. Strukturen, vereinzelte, aber das Kommunikationsziel in keiner Weise beeinträchtigende Fehler                       | beständig hohes Maß an gramm., syntakt. und orthogr. Korrektheit   |
| gut               | Text weitgehend flüssig lesbar u. gut strukturiert; Gedankengang trotz vereinzelter Verzögerungen der Rezeption nachvollziehbar; gewünschte Wirkung erzielt                  | Text wird Aufgabenstellung weitgehend gerecht. Lt. Aufgabenstellung implizit enthaltene od. explizit vorgegebene Punkte werden in der Mehrheit angemessen behandelt     | variabler und präziser Gebrauch, der Textsorte angemessene syntakt. Strukturen, mehrere, aber das Kommunikationsziel in keiner Weise beeinträchtigende Fehler                           | gute Beherrschung der Grammatik, gelegentliche Ausrutscher oder nicht-systematische Fehler, geringfügige Mängel in der Syntax          |
| befriedigend      | Text muss an manchen Stellen mehrfach gelesen werden; Gedankengang mit Brüchen, aber für kooperativen Leser nachvollziehbar; gewünschte Wirkung im Großen und Ganzen erzielt | Text wird Aufgabenstellung im Wesentlichen gerecht. Lt. Aufgabenstellung implizit enthaltene oder explizit vorgegebene Punkte werden nur zum Teil angemessen behandelt  | teilweise eingeschränkter u. nicht immer korrekter, präziser bzw. angemessener Gebrauch; störende (Häufigkeit oder Schwere), aber das Kommunikationsziel nicht beeinträchtigende Fehler | häufige, z. T. systematische gramm. Regelverstöße, die Textrezeption u. Lesefluss beeinträchtigen, wobei Text noch verständlich bleibt |
| ausreichend       | Text muss an vielen Stellen mehrfach gelesen werden; Gedankengang ohne klare Struktur u. nur zum Teil nachvollziehbar; Wirkung/Aussage für Leser nicht klar                  | Text wird Aufgabenstellung teilweise nicht mehr gerecht. Lt. Aufgabenstellung implizit enthaltene od. explizit vorgegebene Punkte nur in Ansätzen angemessen behandelt. | eingeschränkter und/oder nicht korrekter, häufig nicht angemessener Gebrauch; erheblich störende, das Kommunikationsziel beeinträchtigende Fehler                                       | viele systematische, teils elementare Fehler, teilweise oder ganz unverständliche Textpassagen   |
| nicht ausreichend | Text völlig unverständlich; Gedankengang nicht nachvollziehbar, da ohne Struktur; Wirkung: negativ oder gleich null  | Text wird Aufgabenstellung nicht gerecht. Auftreten irrelevanter Inhalte wegen Nicht-Verstehens d. Aufgabenstellung   | in keiner Hinsicht variabel und angemessen  | gramm., syntakt. u. orthogr. Korrektheit unzureichend, Text unverständlich   |
| Faktor            | 2  | 2   | 3   | 3  |
| Teilnote          |  |   |   |  |
| Ergebnis          |  |   |   |  |

Für die einzelnen Bewertungskriterien können die Noten 1 – 5 vergeben werden. Diese sind mit dem entsprechenden Faktor zu multiplizieren und bilden in der Summe einen Punktwert, dem eine Gesamtnote in folgender Weise entspricht:

bis 10 = 1,0

bis 13 = 1,3

bis 17 = 1,7

bis 20 = 2,0

bis 23 = 2,3

bis 27 = 2,7

bis 30 = 3,0

bis 33 = 3,3

bis 37 = 3,7

bis 40 = 4,0

## Corrección y evaluación de pruebas de producción escrita a partir de exámenes de ejemplo

### El examen

Evaluar es un acto con consecuencias. Las notas recibidas pueden alterar positiva o negativamente una nota media, dar o quitar lustre a un certificado o – en el peor de los casos – condenar a un alumno a repetir un curso o un examen. Evaluar conlleva siempre una responsabilidad, tanto en relación al alumno como con nosotros mismos como profesionales, por lo que esta tarea debe realizarse de la forma más objetiva, fiable y válida posible.

Antes de concentrarnos en cómo es posible hacerlo debemos aclarar al menos brevemente estos tres conceptos: Un examen *objetivo* es aquel cuyo resultado no depende del examinador; un examen *fiable* el que no se ve alterado por factores externos, y un examen es *válido* cuando mide (solo) lo que tiene que medir. Estos son los criterios de calidad de un examen más extendidos pero no los únicos. Silvia Prati (2007), por ejemplo, añade la *autenticidad* (existencia de una correspondencia entre la realidad y el examen) y la *factibilidad* (adecuación a las condiciones y a la situación dada) a los ya mencionados.

A continuación se exponen algunas maneras de lograr que una prueba de producción escrita cumpla con estos tres criterios:

- Para que sea fiable no basta con eliminar o reducir el impacto de factores circunstanciales como ruidos, uso de materiales no permitidos, etc., sino que también debe ponerse mucho cuidado en que las instrucciones sean claras y sin ambigüedades, e informar (al menos de forma general) sobre los criterios de evaluación.
- Para que sea válida debe medir exclusivamente la expresión escrita – suponiendo, naturalmente, que nuestra intención es evaluar sólo esta competencia. El concepto de validez se refiere también a la adecuación adecuada al nivel y a los contenidos del curso. Para ello podemos recurrir a los descriptores del MCER. Si, por ejemplo, mis alumnos de nivel B1 deben narrar una historia, puedo consultar en el MCER de qué deben ser capaces a nivel en el tipo de texto “escritura creativa”, y ajustar las

exigencias de la prueba y mis criterios de evaluación a ese nivel de competencia. De todos modos no debemos olvidar que el MCER es un “documento de apoyo”, y que “no establece ningún procedimiento para la elaboración de pruebas de evaluación”. (Soler Canela, 2006)

- Para que sea objetiva podemos aplicar medidas como recurrir a examinadores ajenos a la institución, o impedir que los examinadores sepan de quién es el examen que están corrigiendo (por ejemplo usando números de matrícula en lugar de nombres). Otro método muy extendido para asegurar la objetividad es la corrección por parte de dos o más examinadores. Para asegurar un nivel de objetividad aún mayor cada uno de los examinadores puede poner su nota ignorando la del otro, evitando así que la nota del primer examinador pueda influir en la del segundo.

El método más importante para asegurar la objetividad es, sin embargo, la elaboración de claves de corrección a las que debe ajustarse el examinador durante la corrección. Los examinadores que emplean las mismas claves, al ajustarse a ellas, se ven obligados a evaluar los mismos aspectos de la misma forma, lo que reduce inevitablemente la disparidad de criterio.

### **Las claves de corrección**

Existen muchos modelos de claves de corrección. La elección de uno u otro depende fundamentalmente del enfoque con el que vaya a corregir después: ¿Va a primar la corrección lingüística? ¿La impresión general? ¿O se van a analizar independientemente diferentes aspectos del texto?

#### a) Corrección basada en la corrección lingüística

Este tipo de claves se basan en el recuento del número de errores cometidos. Este total sirve de base para calcular la nota, ya sea mediante un cociente de errores que tiene en cuenta el número de palabras, o bien descontando puntos de un total de puntos. El principal argumento a favor de este enfoque es su altísimo grado de objetividad. Una crítica recurrente es que, al prestar tanta atención a lo que los alumnos hacen mal, pueden contribuir a desmotivarlos. Otras críticas que se suelen hacer a estas claves de corrección fueron rebatidas durante el taller. Por ejemplo no es cierto que penalice más a los alumnos que *arriesgan*, ya que tiene en cuenta la extensión del texto. Otro aspecto crítico (“no es tan matemático como pretende, ya que la decisión de si un determinado error es grave o leve es a menudo subjetiva”) puede contrarrestarse decidiendo previamente qué se considerará leve y grave. A estas claves de corrección suele objetársele también que no se adecuan a la valoración de aspectos como el

estilo o la adecuación del registro, pero es posible compensar esta carencia empleando para ello otras escalas más apropiadas y utilizando el cociente de errores para evaluar la corrección lingüística.

b) Corrección basada en la impresión general (enfoque holístico)

Se trata de un enfoque diametralmente opuesto al de la corrección lingüística: En lugar de cálculos aritméticos lo determinante es la impresión general. Un argumento a favor de este enfoque es que valora un texto tal como la hacen los nativos. El principal argumento en contra es, naturalmente, el elevado riesgo de subjetividad que conlleva. La Association of Language Testers of Europe (2007) recomienda por ello el empleo de descriptores y que haya más de un corrector, para reducir así la subjetividad inherente a este tipo de corrección.

C) Corrección basada en el análisis independiente de diferentes aspectos del texto (enfoque analítico)

Básicamente consiste en asignar puntos de una escala a diferentes aspectos del texto (léxico, ortografía, estilo, estructura, contenido, etc.) y calcular la nota final en base al total de puntos obtenidos. El enfoque analítico se fija por tanto en más aspectos que en la corrección lingüística pero a cambio, pero no excluye la subjetividad en igual medida. Las escalas con las que se valora cada aspecto del texto pueden incluir o no descriptores, como en el ejemplo:

| <b>Corrección lingüística</b> |  |
|-------------------------------|--|
| 4 puntos                      | El texto no tiene errores a nivel lingüístico / El texto apenas tiene errores a nivel sintáctico, morfológico y ortográfico, y éstos no dificultan en absoluto la comprensión del texto.<br>El vocabulario es variado. |
| 3 puntos                      | El texto tiene algunos errores a nivel sintáctico, morfológico y ortográfico que apenas dificultan la comprensión del texto.<br>El vocabulario es variado.   |
| 2 puntos                      | El texto tiene varios errores a nivel sintáctico, morfológico y ortográfico que dificultan la comprensión del texto, aunque (con alguna excepción) no la impiden.<br>El vocabulario es sencillo, pero adecuado.        |
| 0 punto                       | El texto tiene tantos errores a nivel sintáctico, morfológico y ortográfico que impiden o dificultan enormemente la comprensión del texto.   |

Esta escala sirve para ilustrar uno de los problemas típicos en el uso de descriptores que se comentó en el taller: la mezcla de dos o más aspectos evaluables dentro de un mismo descriptor. Por ejemplo: ¿Cuántos puntos debemos dar a un alumno cuyo vocabulario es

“sencillo, pero adecuado” y cuyo texto no contiene “apenas errores a nivel sintáctico, morfológico y ortográfico”? ¿Dos o cuatro?

La siguiente escala, sin descriptores, presenta un problema diferente:

|  |   |   |   |   |   |
|--|---|---|---|---|---|
| Logro de los objetivos comunicativos de la actividad | 0 | 1 | 2 | 3 | 4 |
| Contenido  | 0 | 1 | 2 | 3 | 4 |
| Organización del texto                               | 0 | 1 | 2 | 3 | 4 |
| Estilo   | 0 | 1 | 2 | 3 | 4 |
| Léxico   | 0 | 1 | 2 | 3 | 4 |
| Corrección gramatical                                | 0 | 1 | 2 | 3 | 4 |
| Ortografía   | 0 | 1 | 2 | 3 | 4 |

0: insuficiente, 4: excelente

Puntos totales: \_\_\_\_\_ / 28

Para simplificar el diseño de la tabla, todos los apartados utilizan la misma escala (0-4 puntos), por lo que todos los aspectos a evaluar, independientemente de su importancia, tienen el mismo peso en la nota final. En este ejemplo la ortografía tiene tanta importancia como la corrección gramatical o el léxico. Otro problema que nos encontraríamos al utilizar esta tabla es la dificultad de evaluar por separado los diferentes aspectos. No sólo debido a la existencia de determinados errores de difícil clasificación, sino también a que los diferentes aspectos se influyen mutuamente entre sí: La riqueza léxica y la corrección gramatical influyen en gran medida en el estilo, y la adecuación de éste a la situación comunicativa planteada en el ejercicio determina por su parte si se logran o no los objetivos comunicativos de la actividad.

Estas observaciones no invalidan en modo alguno el enfoque analítico; tan sólo pretenden advertir de los problemas que podemos encontrarnos al aplicarlo.

### **La corrección**

Durante el taller se corrigieron en grupo textos de diferentes niveles tomados de diferentes tipos de exámenes: la parte de producción escrita de un examen de UNICert® III (C1), dos ejercicios de redacción incluidos en exámenes de final de curso (A2 y B1) y un ejercicio de comentario de texto (B2). Los exámenes fueron corregidos a la vez por varios grupos que empleaban diferentes claves de corrección. Los objetivos de esta corrección múltiple eran, primeramente, comparar los resultados y observar si con algún tipo de enfoque se obtenían más notas altas o bajas que con los otros y, además, ver qué problemas concretos surgen en la práctica.

Por lo que respecta al primer punto: Aunque naturalmente no todas las notas fueron iguales, en general las diferencias no fueron grandes, y – lo más importante – no se pudo constatar ninguna relación enfoque/nota del tipo descrito.

A continuación expondré brevemente los principales problemas que aparecieron durante la corrección, y las sugerencias que se hicieron sobre cómo resolverlos:

- El cociente de errores es una buena herramienta para medir la corrección lingüística, pero no para otros aspectos. Los grupos que optaron por emplear este método lo hicieron en combinación con otro tipo de escalas (fundamentalmente analíticas) para compensar sus carencias.
- La valoración del estilo, independientemente del enfoque, es subjetiva, por lo que se sugirió desglosar este aspecto en otros más concretos como, por ejemplo, variedad de vocabulario, uso de las conjunciones, etc.
- Las escalas analíticas no permiten valorar elementos *extra* como la creatividad o la originalidad. Incluir una escala para estos aspectos tampoco sería una buena solución, ya que indirectamente castigaría a las personas poco creativas u originales, que no obtendrían puntos en este apartado.
- Los exámenes con destrezas integradas, en los que la producción escrita se basa en el contenido de un texto presentan un problema específico: ¿Qué hacer cuando el alumno copia frases enteras del texto original? Una solución que encontró bastante aceptación es ignorarlos al leer el texto, hacer como si no existieran, lo que repercutirá negativamente de varias maneras en la nota (menos puntos en contenido y posible penalización por no llegar al número mínimo de palabras). Otra alternativa es la llamada “mediación”: El texto que los alumnos deben resumir o comentar está en su lengua materna, de manera que no sean posibles los *préstamos*. El problema que puede surgir con este método es que los estudiantes extranjeros que hagan un curso de ELE y no dominen el alemán están en inferioridad de condiciones. El poco dominio del alemán sería en este caso un factor que disminuiría la fiabilidad del examen.
- Ninguna escala, ya sea analítica o de coeficiente de errores, sirve para todos los niveles. Ciertos aspectos (como estilo o estructura) resultan difícilmente evaluables en un texto breve de nivel A1 y A2, y ciertos parámetros que pueden tener sentido en estos niveles (“Claridad de expresión”) son superfluos en niveles superiores, ya que la capacidad de expresarse de manera comprensible ya ha sido adquirida hace mucho tiempo. La solución a este problema consiste, obviamente, en la confección de claves

de corrección específicas para cada uno de los niveles. Si el mismo problema puede darse en diferentes pruebas del mismo nivel, se debe adaptar la clave de corrección a la prueba concreta que se desea evaluar.

## **Bibliografía**

PRATI, Silvia (2007): *La evaluación en español lengua extranjera: La elaboración de exámenes*. Libros de la Araucaria: Buenos Aires.

SOLER CANELA, Oscar (2006): *La evaluación de ELE y el Marco Común Europeo de Referencia*. redELE [en línea]

<http://www.mepsyd.es/redele/Biblioteca2006/angep/ComunicacionSoler.pdf> [Consulta: 29/09/2009]

VARIOS AUTORES (2007): *ALTE-Handreichungen für Testautoren. Modul 3-4*. ALTE (Association of Language Testers of Europe) [en línea]

<http://www.alte.org/downloads/index.php> [Consulta: 29/09/2009]

VARIOS AUTORES (2002): *Marco Europeo de Referencia para las Lenguas: Aprendizaje, enseñanza, evaluación*. Anaya: Madrid / Instituto Cervantes [en línea]

[http://cvc.cervantes.es/obref/marco/cvc\\_mer.pdf](http://cvc.cervantes.es/obref/marco/cvc_mer.pdf) [Consulta: 28/09/2009]

VARIOS AUTORES (2006): *Plan curricular del Instituto Cervantes: Niveles de referencia para el español*. Biblioteca Nueva: Madrid.



## **LESEVERSTEHEN: OFFENE PRÜFUNGSAUFGABEN UND DEREN BEWERTUNG AUF DEN STUFEN UNICert® II UND III**

**Teresa Eichelmann, TU Dresden  
Alexander Biewald, Universität Innsbruck**

### **1. Einleitung**

Als eine der vier kommunikativen Sprachhandlungen ist die (Weiter-) Entwicklung des *Leseverstehens* Bestandteil einer jeden Fremdsprachenausbildung. Die Überprüfung dieses *skills* erfolgt jedoch auf vielfältige Weise, obgleich im hochschulspezifischen UNICert®-Kontext zunehmend *handlungsorientierte Ansätze* verwendet werden. Den Prüfungen werden dabei realitätsnahe Kommunikationssituationen zu Grunde gelegt, die meist anhand von *offenen Aufgabenstellungen* bearbeitet werden. Im Hinblick auf die Überprüfung des Leseverstehens werden derartige Aufgaben allerdings besonders stark diskutiert, da sie nur in Verbindung mit produktiven Fähigkeiten und Fertigkeiten *bewältigt* und *bewertet* werden können.

Anhand der skizzierten Kontroverse stellt sich daher die Frage, wie offene, handlungsorientierte Prüfungsaufgaben für das Leseverstehen unter Einbeziehen *skill-übergreifender Konzeptionen* gestaltet werden sollten. Damit einher geht auch die Problematik um die diffizile Bewertung entsprechender Aufgabenstellungen, wobei ein möglicher Lösungsansatz in der *Spezifikation von Bewertungskriterien* besteht.

Ziel der AG 3/Englisch war es daher, mit Hilfe von ausgewählten Texten *handlungsorientierte, offene Aufgaben* zur Überprüfung des *Leseverstehens* auf den *Niveaustufen UNICert® II und III* zu erstellen sowie *Kriterien für deren Bewertung* zu erarbeiten.

### **2. Theoretische Grundlagen**

Die Vorbereitung auf das Entwickeln von Prüfungsaufgaben und Bewertungskriterien umfasste zunächst die Wiederholung grundlegender Aspekte des Testens kommunikativer Kompetenzen. Dabei wurden die Quantität und Qualität der für das Leseverstehen relevanten Sprachhandlungen beschrieben und in Abhängigkeit von den Niveaustufen UNICert® II und III differenziert. Ergänzend dazu konnten Prüfungsformate und deren Charakteristika resümiert und im Hinblick auf ihre Angemessenheit im UNICert®- Kontext besprochen werden, wobei die Qualitätskriterien Validität, Reliabilität und Objektivität als Merkmale einer jeden Prüfungsaufgabe stark diskutiert wurden. Zudem definierten die Teilnehmer zwei Verfahren zur Bewertung von Prüfungsleistungen und erörterten diese in Bezug auf ihre Vor- und Nachteile sowie hinsichtlich ihre Eignung im Hochschulkontext.

## 2.1 Leseverstehen: Differenzierung von Qualität und Quantität der Sprachhandlungen auf den Niveaustufen UNICert® II und III

Den Anfang der AG- Arbeit bildete die Auseinandersetzung mit den Deskriptoren des Gemeinsamen europäischen Referenzrahmens (GeR), die das Leseverstehen als rezeptive kommunikative Aktivität beschreiben. Unter Berücksichtigung der avisierten Niveaustufen B2/UNICert® II und C1/III wurden dabei 11 Deskriptoren<sup>1</sup> mit dem Ziel bearbeitet, die beim Leseverstehen vollzogenen Sprachhandlungen zu bestimmen und den Ausprägungsgrad dieser Fähigkeiten und Fertigkeiten zu verdeutlichen. Tabelle 1 fasst die Ergebnisse zusammen.

|  | B2/UNICert® II und C1/UNICert® III  |
|--|---|
| <b>Sprachhandlung</b>                  | <ul style="list-style-type: none"> <li>• <i>lesen und verstehen</i> <ul style="list-style-type: none"> <li>○ globales Verstehen B2</li> <li>○ detailliertes Verstehen C1 bzw. B2 mit Hilfsmitteln</li> </ul> </li> </ul>  |
| <b>Texte und deren Charakteristika</b> | <ul style="list-style-type: none"> <li>• <i>lange, komplexe Texte</i> <ul style="list-style-type: none"> <li>○ <i>innerhalb</i> des eigenen Fach- und Interessengebietes B2</li> <li>○ <i>außerhalb</i> des eigenen Fach- und Interessengebietes C1 bzw. B2 mit Hilfsmitteln</li> </ul> </li> </ul> |
| <b>Hilfsmittel</b>                     | <ul style="list-style-type: none"> <li>• Frequenz der Nutzung von Hilfsmitteln</li> <li>• Zuhilfenahme des Wörterbuchs</li> <li>• mehrfaches Lesen schwieriger Passagen</li> </ul>  |

- Tabelle 1 -

Die Betrachtung der Deskriptoren ermöglichte es, die Kompetenzprofile der Niveaustufen B2 und C1 zu spezifizieren und mit dem Raster (Tabelle 1) ein Instrument zu erstellen, das zur besseren Unterscheidung der Kompetenzstufen herangezogen werden kann.

Trotz der erfolgreichen Systematisierung wurden auch problematische Aspekte des Referenzrahmens diskutiert, beispielsweise die Interpretation von mit vagen Formulierungen durchsetzten Deskriptoren. Die Teilnehmer beschrieben dies als sehr schwierig und kritisierten die damit einhergehende hohe Subjektivität. Als Voraussetzung für die Erhöhung der Objektivität wurde die Ausbildung eines gemeinsamen Verständnisses der Kompetenzbeschreibungen genannt, welches – in erster Näherung – auf Grundlage oben beschriebener Orientierungshilfen erzielt werden kann.

## 2.2 Qualitätskriterien

In einem weiteren Arbeitsschritt wurden die Qualitätskriterien *Validität* und *Reliabilität* sowie die bereits zuvor diskutierte *Objektivität* definiert und auf das folgende Prüfungsbeispiel angewendet:

<sup>1</sup> vgl. Anhang 4.1

**Reading 1: International Polar Year: In from the cold (nature news, 25 Feb. 2009)**

A/\_\_\_\_\_ Scientists need such sophisticated equipment in order to tackle serious problems. Most importantly, this IPY has proceeded amid galloping climate change, which has already altered the Arctic in dramatic ways and is increasingly making itself apparent around Antarctica. The scale of the polar changes dwarfs what many researchers would have thought possible even just a few years ago.

B/\_\_\_\_\_ It might therefore appear that, as so often in the past, science reigns supreme at the planet's poles. But as climate change opens up vast parts of the Arctic to commerce, nations are starting to exert their influence in the region more purposefully, and political tensions that have been building up over the last decades might soon explode.

C/\_\_\_\_\_ The fourth International Polar Year (IPY), which ran from March 2007 to March 2009, has seen some 50,000 scientists, students, technicians and crew participate in expeditions and scientific programmes focusing on the Arctic and Antarctic. Its 170 collaborative international projects have covered disciplines from ecology to astronomy, from the social sciences to human and animal health. In the Arctic, indigenous people and their traditional knowledge became part of the science wherever possible, sometimes inspiring it in unanticipated ways.

D/\_\_\_\_\_ For this reason, the current IPY, which extends a tradition that reaches back almost 130 years, takes a more comprehensive approach. In doing so, however, it is also true to tradition. Each of the three previous polar years – in 1882-83, 1932-33 and 1957-58 – kicked off a new era in polar science. In the most recent of these, the main scientific challenge was to explore the most remote and hostile spots on the globe. The focus over the past two years, by way of contrast, has been on studies that encompass the physical, biological and social components of the polar environments.

E/\_\_\_\_\_ In studying these areas of science, many polar researchers now take a systems approach, examining, for example, interactions between glaciers, ocean circulation, sea ice, algae and seal health. These and many other parts of the polar system have been studied before. But the scope of activities carried out during this IPY would have been unthinkable without the technological advances of recent decades, including such tools as satellites, laptop computers and high-tech clothing.

**Task 1.1: Re-establish textual order by numbering each of the paragraphs to form a coherent text. (1 pt per answer) \_\_\_\_\_ / 5 pts**

**Task 1.2: Answer the questions briefly with information from the text. (2 pts per answer) \_\_\_\_\_ / 8 pts**

1. What would many researchers not have thought imaginable a few years ago?
2. What seems to be the status of science as regards discussions on the situation in the Arctic?
3. What characterises the projects that have taken place in the context of the 4<sup>th</sup> IPY?
4. What was the most important undertaking within the 3<sup>rd</sup> IPY?

**Validität:**

Die Validität bzw. *Gültigkeit* gibt an, inwieweit der Test die Kompetenz, die er zu messen versucht, tatsächlich misst. Der Test ist dann *valide*, wenn die Testergebnisse einen

unmittelbaren und fehlerfreien Rückschluss auf den Ausprägungsgrad der zu erfassenden Kompetenz zulassen.

#### *Diskussion Task 1.1 und Task 1.2*

Die Teilnehmer beurteilten die Aufgabe 1.1 zur Wiederherstellung eines kohärenten Textes als valide und stützten ihre Argumentation auf die darin geprüften Kompetenzen. Sowohl die Fähigkeit/Fertigkeit des globalen Verstehens als auch die Kenntnis von Mitteln zur Textstrukturierung sind Komponenten des Leseverstehens, die direkt, also ohne den Einfluss einer produktiven Fähigkeit/Fertigkeit geprüft werden.

Auch die sich anschließende Aufgabe 1.2 wurde als valide beschrieben, da sie m.H. von *short answer questions* das detaillierte Leseverstehen prüft. Kritisch diskutiert wurde allerdings der aus dem Aufgabenformat resultierende Einfluss produktiver *skills*. Um die dadurch geminderte Validität zu erhöhen, wurde die unabhängige Bewertung der inhaltlichen und sprachlichen Richtigkeit der Antwort vorgeschlagen.

#### **Reliabilität:**

Die Reliabilität beschreibt den Grad der *Genauigkeit*, mit dem der Test ein Verhaltensmerkmal abbildet. Der Test ist dann *reliabel*, wenn mit Hilfe der Testergebnisse die Kompetenz(en) des Lernenden genau beschrieben werden können.

#### *Diskussion Task 1.1 und Task 1.2*

Beide Tasks wurden von den Teilnehmenden als nur mäßig reliabel bewertet. Eine Ursache dafür war das in 1.1 verwendete Aufgabenformat, welches voneinander abhängige Items beinhaltet, die bei falscher Positionierung eines Abschnitts (= Items) unvermeidbare Folgefehler bzgl. der Anordnung weiterer Paragraphen verursachen. Um diese dennoch valide Aufgabe beibehalten zu können, wurde ein erweitertes Bewertungsverfahren besprochen, das sowohl die Position als auch die Verknüpfung der Abschnitte berücksichtigt.

Im Vergleich dazu weist Task 1.2 vier unabhängige Fragen auf, deren Validität unglücklicherweise durch die wenig genaue Aufgabenstellung „...with *information from the text*.“ limitiert wird. Diese könnte jedoch durch Umformulieren oder das Hinzufügen von „Do not quote.“ präzisiert werden, womit auch ein Abschreiben des Textes unterbunden würde. Ferner scheint die Einleitung der Fragen mit Hilfe des „what“ problematisch, da dies u.U. ein breites Spektrum möglicher Antworten forciert, deren Bewertung diffizil wäre. Um die Eindeutigkeit der Fragen zu erhöhen, könnten Operatoren wie „name“ oder „list“ eingesetzt werden.

#### **Objektivität:**

Die Objektivität beschreibt den Grad der *Unabhängigkeit des Testergebnisses* von jeglichen Einflüssen außerhalb der getesteten Person.

#### *Diskussion Task 1.1 und Task 1.2*

Aufgabe 1.1 wurde zunächst als objektiv beurteilt, da es nur *eine* richtige Folge der Textabschnitte gibt. Die Teilnehmer beanstandeten allerdings die damit einhergehende

Abhängigkeit der Paragraphen, die – wie bereits diskutiert – Reliabilität und Objektivität negativ beeinflusst. Um dem entgegenzuwirken und schließlich ein weitgehend objektives Testergebnis zu erzielen, bedarf es präziser Bewertungskriterien, die von allen Prüfern einzuhalten sind.

Als äußerst komplex erwies sich die Diskussion der Objektivität von Task 1.2, da diese Aufgabe nur durch das gleichzeitige Anwenden rezeptiver und produktiver Fähigkeiten/Fertigkeiten bewältigt werden kann. Erschwerend hinzu kam die offene Aufgabenstellung, deren Bewertung häufig sehr subjektiv ist. Um die Objektivität dennoch zu gewährleisten, befürworteten die Teilnehmer eine getrennte Bewertung der inhaltlichen und sprachlichen Aspekte der Antwort mit Hilfe eines im Voraus entwickelten Erwartungshorizonts. Sprachlich- produktive Leistungen sollten separat von mehreren Prüfern auf Grundlage detaillierter Kriterien gewertet werden, wobei eine abschließende Einigung über das Ergebnis erforderlich ist.

### 2.3 Aufgabenformate zur Überprüfung des Leseverstehens

Im Anschluss an die Diskussion der Qualitätskriterien sammelten und gruppieren die Teilnehmer im Rahmen eines Brainstormings Aufgabenformate zur Überprüfung des Leseverstehens. Offene Formate wurden dabei in besonderem Maße berücksichtigt, da sie zu einem späteren Zeitpunkt bei der Konstruktion von Prüfungsaufgaben für die Niveaustufen UNlcert® II und III angewendet werden sollten. Tabelle 2 fasst die Ergebnisse zusammen.

| offene Formate   | geschlossene Formate  |
|--|---|
| <ul style="list-style-type: none"> <li>• <i>sentence completion</i></li> </ul>   | <ul style="list-style-type: none"> <li>• <i>yes/no</i> bzw. <i>true/false items</i></li> </ul>  |
| <ul style="list-style-type: none"> <li>• <i>short answer items</i>, z.B.:<br/>Paraphrasen</li> </ul>   | <ul style="list-style-type: none"> <li>• <i>(multiple) matching</i>, z.B.:<br/>Zuordnung von Überschriften zu Textabschnitten</li> </ul>                    |
| <ul style="list-style-type: none"> <li>• <i>open- ended questions</i>, z.B.:<br/>„W“- Fragen</li> </ul>  | <ul style="list-style-type: none"> <li>• <i>information transfer</i>, z.B.:<br/>Beschriftung einer Abbildung m.H. der Informationen aus dem Text</li> </ul> |
| <ul style="list-style-type: none"> <li>• Auflistung von <i>pros and cons</i></li> </ul>  | <ul style="list-style-type: none"> <li>• <i>cloze test</i> bzw. <i>c- test</i></li> </ul>   |
| <ul style="list-style-type: none"> <li>• multiple choice:<br/>geschlossen, wenn Antwortmöglichkeiten vorgegeben sind<br/>offen, wenn Antwortmöglichkeiten selbst erstellt werden müssen</li> </ul> |   |

- Tabelle 2 -

Für den hochschulspezifischen Kontext sind offene Aufgabenformate von großer Bedeutung, da sie ein integratives Testen u.U. sogar produktiver und rezeptiver Fähigkeiten und Fertigkeiten ermöglichen. Dabei ähnelt die zur Bewältigung einer offenen Aufgabe häufig

notwendige kombinierte Anwendung verschiedener skills dem kommunikativen Verhalten in einer realen Situation. Dieses Potenzial ist Grundlage der Konzeption von in einen – meist fachbezogenen – Kontext eingebetteten, handlungsorientierten UNICert®- Prüfungen.

Während der Diskussion in der Arbeitsgruppe wurde jedoch ein noch großes Unbehagen gegenüber der Anwendung entsprechender Aufgabenstellungen deutlich. Ursache dafür mag die anhaltende Dominanz geschlossener Formate sein, die nicht zuletzt auf deren positive Eigenschaften (vgl. **2.2 Qualitätskriterien**) zurückzuführen ist. Damit verbunden ist auch die möglicherweise mangelnde Erfahrung hinsichtlich der Konstruktion offener Formate, die die auffallende Präferenz für die Gestaltung geschlossener Aufgaben weiter unterstützt.

## **2.4 Bewertungsverfahren**

Mit dem Ziel, die Besprechung der theoretischen Grundlagen von Testtheorie und Testkonstruktion zu beenden, wurden schließlich zwei Verfahren zur Bewertung von Prüfungsaufgaben erörtert.

Das *analytische Vorgehen* beschreibt die separate Bewertung aller Teilaspekte einer Aufgabe, die schließlich in einem Gesamturteil zusammengeführt werden. Voraussetzung dafür ist die Identifikation und qualitative Beschreibung jener zur Bewältigung der Aufgabe notwendigen Fähigkeiten/Fertigkeiten, die – der Beurteilung zu Grunde gelegt – die Objektivität verbessern. Problematisch erscheint jedoch die geringe Validität dieses reliablen, obgleich sehr zeitaufwändigen Verfahrens. Ursache dafür ist die wenig reale, zu Bewertungszwecken vorgenommene Trennung der skills, die u.U. einen verzerrten Gesamteindruck über die erbrachte Leistung verursachen kann.

Im Gegensatz dazu wird beim *holistischen Vorgehen* die Prüfungsaufgabe auf Basis dieses Gesamteindrucks bewertet. Ein Vorteil des Verfahrens ist der dafür notwendige geringe Zeitaufwand, der eine multiple Bewertung durch verschiedene Prüfer gestattet und somit die scheinbare Subjektivität jeder Einzelbewertung minimiert. Eine holistische Bewertung sollte allerdings ausschließlich von erfahrenen Prüfern durchgeführt werden.

Abschließend diskutierten die Teilnehmer die Eignung beider Bewertungsverfahren im UNICert®- Kontext, wobei das analytische Vorgehen auf Grund der hohen Genauigkeit der Beurteilung favorisiert wurde. Es ist damit nicht nur möglich, die Subjektivität der Bewertung zu mindern, sondern auch den Lernenden ein konkretes Feedback über ihre Leistung zu geben. Trotz dieser Vorteile scheint die Kombination beider Verfahren empfehlenswert, da mit Hilfe des holistischen Gesamteindrucks die analytisch ermittelte Leistung bestätigt werden kann.

## **2.5 Entwicklung von Prüfungsaufgaben mit Bewertungskriterien für das Leseverstehen: Ergebnisse**

Die Entwicklung einer Prüfungsaufgabe für das Leseverstehen erfolgte auf Basis der folgenden Aufgabenstellung:

*„Entwickeln Sie eine skill-übergreifende, offene und handlungsorientierte Aufgabe zur Überprüfung des Leseverstehens und spezifizieren Sie dazu Bewertungskriterien. Verwenden Sie als Basis den Fachartikel „Genetically Modified Foods“.*

### 2.5.1 Gedanken zur Prüfungsentwicklung

Um diese Prüfung *skill- übergreifend* zu gestalten, wird das Leseverstehen mit den produktiven Fähigkeiten/Fertigkeiten Schreiben bzw. Sprechen kombiniert. Der Artikel „Genetically Modified Foods“ dient dabei als Basis, sprich Impulstext, für die Gestaltung der Prüfungsaufgaben, für deren Bearbeitung die Prüfungskandidaten nicht nur beweisen müssen, dass sie den Text grundsätzlich verstehen, sondern auch, dass sie die darin enthaltenen Informationen für die selbständige Produktion weiter verwenden können.

Um die Prüfung *handlungsorientiert* zu gestalten, wurden zwei Situationen entworfen, in denen eine Fachkraft „gezwungen“ wäre, sich mit dem Artikel genauer auseinanderzusetzen. Diese Situationen – als Problem definiert – sollen die Kandidaten in eine konkrete Rolle hineinversetzen und der Handlung eine gewisse Dringlichkeit („exigency“) verleihen.

Zu diesem Zweck wurden einige kurze Aufgaben (siehe Aufgaben A bis C unten) erstellt, die den Studierenden erlauben, sich genauer mit dem Text und dessen Inhalt auseinanderzusetzen. Diese Aufgaben sind normale, eher geschlossene Leseverstehensübungen, die dennoch so konzipiert sind, dass sie die Prüfungskandidaten zum mündlichen bzw. schriftlichen Produktionsteil führen. Der Produktionsteil ist die eigentliche *offene Aufgabe*, in der die Kandidaten die Impulse des Artikels mit ihrem Fachwissen aus dem Studium verbinden können, um ein detailliertes Schriftstück bzw. eine überzeugende Präsentation zu verfassen.

N.B.: Diese Prüfungsaufgaben sind für das **Niveau C1/Englisch** konzipiert. Natürlich wäre die ganze Prüfung – auch die Anweisungen – auf Englisch. Innerhalb des Workshops wurde allerdings deutsch gesprochen, sodass die Aufgaben zunächst auch in dieser Sprache erarbeitet wurden.

### 2.5.2 Prüfungsbeispiel

Liebe Prüfungskandidatin, lieber Prüfungskandidat!

Um den verschiedenen Prüfungsaufgaben ein professionelles Setting zu geben, sollten Sie sich in **eine** der zwei unten stehenden Rollen hineinversetzen. Die kurze Szenariobeschreibung (sprich: Problemdefinition) soll Ihnen helfen, die aktuelle Lage und Perspektive der Person zu verstehen und zu verinnerlichen, um während der Prüfung realitätsnah denken und agieren zu können.

**1. PR Manager/-in bei Monsanto Europe S.A.:** Als PR Manager/-in ist es Ihre Aufgabe, die Firma Monsanto nach außen stets positiv zu repräsentieren. Aufgrund negativer Presseberichte zum Thema „*Genmanipulierte Lebensmittel und die Rolle der Großkonzerne*“ leidet das Image von Monsanto allerdings erheblich (ersichtlich unter anderem durch einen fallenden Aktienkurs). Diese mediale Hetzkampagne beruht offenbar nicht auf wissenschaftlichen Erkenntnissen und schürt die Angst in der Bevölkerung.

**2. Ökobauer:** Als Bauer haben Sie eigentlich schon immer naturtreu gearbeitet, wobei dies auch für viele andere Bauern in Ihrer Region galt. Vor 10 Jahren wurden Sie als Ökobauer zertifiziert und einige Nachbarn sind Ihrem Vorbild gefolgt. Mehrere Landwirte der Region haben seitdem allerdings die wesentlich lukrativeren Verträge der Großkonzerne angenommen und bauen deshalb nun genmanipulierte Nutzpflanzen an. Diese „moderne“

Landwirtschaft ist zu einer echten Bedrohung geworden, nicht nur für die „umzingelten“ Ökobauern sondern auch für das Ökosystems der gesamten Region selbst.

### **Aufgabe A: Definitionen**

Finden Sie die Wörter im Text (in den ersten 3 Absätzen), die zu folgenden Definitionen passen. Tragen Sie die Wörter in die Tabelle ein. (20 Punkte)

|  |                   |
|--|-------------------|
| 1. a person who espouses or supports a cause, etc. (n.)            | proponent         |
| 2. to give, grant, or bestow (v.)                                  | confer            |
| 3. a reserve supply of goods stored (n.)                           | stockpile         |
| 4. to die (v.)   | succumb           |
| 5. the act of growing agricultural products (n.)                   | cultivation       |
| 6. reason for disapproving, or disliking (n.)                      | objection         |
| 7. the act make use of something mainly for profit (n.)            | commercialisation |
| 8. to improve the quality or condition of (v.)                     | enhance           |
| 9. to help to grow or develop; stimulate; promote                  | foster            |
| 10. The amount produced; the return on labor, investment, etc. (n) | yield             |

### **Bewertungskriterien:**

In Aufgabe A werden für jede richtige Antwort 2 Punkte vergeben. Weil die Aufgabe reines Leseverstehen prüft, wird die Rechtschreibung bzw. Verwandlung in die Grundform nicht bewertet (z.B. ein Kandidat könnte das Grundwort „confer“ mit doppel „r“ schreiben (im Text steht „conferring“)).

### **Aufgabe B: Fragen zum Text**

Beantworten Sie folgende Frage an Hand der Informationen im Artikel. Bitte achten Sie darauf, dass Sie den Sinn so weit wie möglich mit Ihren Worten umschreiben. (4 Punkte)

1. On what basis can European countries reject genetically modified plants?

*As a part of the legislation for the use of GM crops, the European Commission formulated a law that allows individual national governments to ban specific GM crops if they find new evidence of risks connected with cultivating and consuming them.*



**Bewertungskriterien:**

In Aufgabe B werden für jede richtige Antwort 4 Punkte vergeben. Ein Problem bei der Bewertung ist die vage Bedeutung von „umschreiben“: muss jedes Wort aus den Text umschrieben werden oder nur manche? Wenn der Kandidat in seiner Antwort den Begriff „legislative provision“ (siehe Artikel, 4. Absatz), kann der Prüfer dann sicher sein, dass der Kandidat weiß, was eine „provision“ im Englischen ist?

**Aufgabe C: Argumente für und gegen**

Tragen Sie je fünf Argumente für bzw. gegen die Verwendung von genetisch modifizierten Nutzpflanzen in die Tabelle ein. (Es ist möglich, dass der Text darüber hinaus weitere Argumente enthält- diese müssen Sie nicht eintragen). (40 Punkte)

| Arguments that speak <b>for</b> the genetic modification of crops  | Arguments that speak <b>against</b> the genetic modification of crops  |
|--|--|
| <ul style="list-style-type: none"> <li>Genetic modification makes crops resistant to pests (insects, fungi, viruses).</li> </ul>   | <ul style="list-style-type: none"> <li>Manipulating the genes is an unnatural process and therefore unnecessary.</li> </ul>  |
| <ul style="list-style-type: none"> <li>Genetic modification makes crops resistant to herbicides, which means that herbicides can be used to kill weeds without harming the crops.</li> </ul> | <ul style="list-style-type: none"> <li>Genetically modified foods are unsafe for human consumption (we don't know how the changed genetics interact with our bodies).</li> </ul> |
| <ul style="list-style-type: none"> <li>Genetic modification allows farmers to produce more crop per hectare.</li> </ul>  | <ul style="list-style-type: none"> <li>Genetic modification harms the environment when the modified genes from crops can "escape" to wild species.</li> </ul>                    |
| <ul style="list-style-type: none"> <li>Genetic modification may give the crop a higher nutritional value.</li> </ul>   | <ul style="list-style-type: none"> <li>Genetic modification harms the environment because resistant crops stimulate the evolution of resistance in pests.</li> </ul>             |
| <ul style="list-style-type: none"> <li>The higher efficiency in farming genetically modified crops should reduce prices for consumers.</li> </ul>  | <ul style="list-style-type: none"> <li>In general, GM foods are unwanted by humans.</li> </ul>   |

### *Bewertungskriterien:*

In Aufgabe C werden für jede richtige Antwort 4 Punkte vergeben. Die Tabelle zwingt den Kandidaten nach 10 Argumente zu suchen. Hierbei ist es notwendig, die verschiedenen Punkte des Autors zu analysieren.

### **Aufgabe D: Produktion**

**Vorschlag 1 (mündliche Produktion):** Als PR Manager/-in von Monsanto haben Sie die Aufgabe, das geschädigte Image der Firma wieder aufzupolieren. Sie entscheiden sich (als erster kleiner Schritt), einen Diskussionsabend an der örtlichen Fachhochschule zu gestalten. In Ihrer prägnanten **Präsentation** werden Sie auf ausgewählte Argumente (max. 3!) näher eingehen, um das kritische Publikum – bestehend aus regionalen Politikern sowie Vertretern der Wirtschaft und Landwirtschaft – von den Vorteilen der GM-Landwirtschaft zu überzeugen.

**Vorschlag 2 (mündliche Produktion):** Als Pionier der Ökolandwirtschaft wurden Sie zum European Farmers' Congress 2010 eingeladen, um **einen Vortrag** zu halten. Sie schätzen, dass ca. 70 % der Teilnehmer (Groß- und Kleinlandwirte; Vertreter aus der Industrie) des Kongresses nicht nach ökologischen Grundsätzen arbeiten. Sie müssen daher mit einer starken Opposition rechnen. Nichtsdestotrotz haben Sie es sich vorgenommen, das Kongresspublikum von den Risiken und Gefahren der GM-Landwirtschaft zu überzeugen. Zu diesem Zweck wählen Sie die stärksten Argumente (max. 3!). Sie können sich sehr gut auf eigene Erfahrungen und Geschichten berufen, denn die ökologischen Anbauflächen in Ihrer Region sind regelrecht von GM-Versuchsfeldern umzingelt.

**Vorschlag 3 (schriftliche Produktion):** Die Geschäftsführung der Firma Monsanto ist an Sie als PR Manager/-in herangetreten, um das Image der Firma wieder aufzupolieren. Sie verlangt von Ihnen **einen Bericht** darüber, wie das Image der Firma bzw. die Akzeptanz von GM-Lebensmitteln verbessert werden könnte.

*NB: Bei **Vorschlag 3** besteht die Gefahr, dass die Prüfungskandidaten/-innen für diese Aufgabe einfach eine Liste von willkürlichen Lösungsansätzen erstellt (grob gesagt: „Wir können das machen... Oder wir können auch jenes machen... Zusätzlich besteht noch die Möglichkeit, etwas anderes zu machen“). So eine Liste ist weder eloquent noch professionell und somit auch unrealistisch. In anderen Worten, es ist nicht die Aufgabe eines hoch bezahlten PR Managers, einfaches Brainstorming zu betreiben. Realistischer wäre, wenn die Prüfungskandidaten/-innen (sprich: der/die PR Manager/-in) **einen besten** Lösungsansatz auswählt und versucht, diesen vor der Geschäftsführung zu verteidigen. Hierin besteht der große Unterschied zwischen informationsorientiertem Denken/Handeln und Argumentation.*

**Vorschlag 4 (schriftliche Produktion):** Als einer von wenigen Ökobauern in Ihrer Region sind Sie von der GM-Landwirtschaft geplagt. Sie entscheiden sich deshalb, einen **Brief an den Redakteur** der regionalen Zeitung zu schreiben, um die Bevölkerung über die fraglichen Entwicklungen in der industriellen Landwirtschaft aufzuklären und sie mittels stichhaltiger Argumente (max. 3!) zu überzeugen, dass die GM-Landwirtschaft unauffällige aber gravierende Gefahren in sich birgt.

### 3. Bibliographie

- Hughes, Arthur (2003). *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- Lienert, Gustav und Ulrich Ratz (1998): *Testaufbau und Testanalyse*. Weinheim: Psychologie Verlags Union.
- Rost, Jürgen (2004): *Lehrbuch Testtheorie- Testkonstruktion*. Bern: Verlag Hans Huber.

### 4. Anhang

#### 4.1 Deskriptoren für das Leseverstehen

##### **UNicert® II/B2**

- Kann **Fachartikel, die über das eigene Gebiet hinausgehen**, **lesen und verstehen**, **wenn er/sie ab und zu im Wörterbuch nachschlagen kann**, um das Verständnis der verwendeten Terminologie **zu überprüfen**.
- Kann **lange, komplexe Anleitungen im eigenen Fachgebiet verstehen**, auch **detaillierte Vorschriften oder Warnungen**, **sofern schwierige Passagen mehrmals gelesen werden können**.
- Kann **rasch den Inhalt und die Wichtigkeit von Nachrichten, Artikeln und Berichten über Themen, die mit meinen Interessen oder meinem Beruf zusammenhängen**, **erfassen und entscheiden, ob sich ein genaueres Lesen lohnt**.
- Kann **Artikel und Berichte über aktuelle Fragen lesen und verstehen**, in denen die **Schreibenden eine bestimmte Haltung oder einen bestimmten Standpunkt vertreten**.
- Kann **Korrespondenz zu Themen des eigenen Fach-, Studien- oder Interessengebietes lesen** und die **wesentlichen Punkte erfassen**.

##### **UNicert® III/C1**

- Kann **lange, komplexe Texte im Detail verstehen**, **auch wenn diese nicht dem eigenen Spezialgebiet angehören**, **sofern schwierige Passagen mehrmals gelesen werden können**.
- Kann **unter gelegentlicher Zuhilfenahme des Wörterbuchs jegliche Korrespondenz verstehen**.
- Kann **ein weites Spektrum langer, komplexer Texte, denen man im gesellschaftlichen und beruflichen Leben sowie in der Ausbildung begegnet**, **verstehen und dabei feinere Nuancen auch von explizit oder implizit angesprochenen Einstellungen und Meinungen erfassen**.
- Kann **lange, komplexe Anleitungen für neue Geräte oder neue Verfahren auch außerhalb des eigenen Fachgebiets im Detail verstehen**, **sofern schwierige Passagen**

mehrmals gelesen werden können.

- Kann hoch spezialisierten Texten aus dem eigenen Fachgebiet Informationen, Gedanken und Meinungen entnehmen.

*Legende:*                    rot markiert = *Texte und deren Charakteristika*  
                                  blau markiert = *Hilfsmittel*  
                                  grün markiert = *Sprachhandlung*

Das Thema der AG 3 lautete: *Erstellen von offenen Prüfungsaufgaben und deren Bewertung (sprachspezifisch)*. Im Einzelnen wurden folgende Punkte behandelt:

1. UNIcert-Kriterien für Prüfungsaufgaben
2. UNIcert-Kriterien für Stufenadäquatheit
3. Fachbezogene Texte und Aufgaben für UNIcert II
4. Studienfachübergreifende Texte und Aufgaben für UNIcert III
5. Fachbezogene Texte und Aufgaben für UNIcert III
6. Kriterien zur Unterscheidung von UNIcert II und UNIcert III
7. Bewertungskriterien

Wir legten als Qualitätskriterien die von der UNIcert®-Kommission erarbeitete *Handreichung zur Bewertung von Prüfungsaufgaben* von 2008 zugrunde. Sie verzeichnet u.a. zum verstehenden Lesen und zur schriftlichen Textproduktion formale Anforderungen an Prüfungen, Kriterien zur Textauswahl sowie zur Aufgabenstellung. Diese für alle vier UNIcert®-Stufen gültigen Leitlinien wurden für die Stufen II und III ergänzt durch die Deskriptoren, die während des internen workshops in Rostock 2007 formuliert wurden. Anhand dieser Kriterien wurden UNIcert-Prüfungen der Stufe II und Stufe III aus dem fachbezogenen Französischunterricht für Studierende der Rechtswissenschaft analysiert.

Als Beispiel für studienfachübergreifende Texte und Aufgaben der Stufe III wurde eine Unterrichtsreihe vorgestellt, die auch in Kursen für BA-Studierende der Galloromanistik eingesetzt werden kann. Alle Texte beziehen sich auf das Rahmenthema: *l'Allemagne avant et après la chute du mur vue par des germanistes français*. Zur Auswahl stehen sozialpsychologische Texte, politikwissenschaftliche Themen, ein Beitrag zu ökonomischen Aspekten und auch Interpretationen literarischer Texte, z.B. zu Jurek Becker: *Amanda herzlos* in französischer Sprache. Die Texte liegen in akustischer (akademische Vorträge) und gedruckter Form vor. Als Prüfungstext wurde ein Artikel aus *Le Monde diplomatique*, August 2004, gewählt: *Quinze ans après la chute du mur de Berlin - Les Allemands de l'Est saisis par l'Ostalgie*. Das Leseverstehen wird anhand von Fragen zum Globalverständnis, zum aufsuchenden Lesen und zu Einzelheiten des Textes überprüft. Zum schriftlichen Ausdruck wurde folgendes Thema gestellt, das in 90 Minuten abzuhandeln war: „Dans une entrevue accordée au journal *Freitag* et publiée le 25 décembre 2000, Lothar de Maizière déclare : « Trop longtemps nous avons été un peuple dans deux Etats différents et nous souffrons maintenant d'être deux peuples dans un seul Etat. » Expliquez, puis discutez ce point de vue en vous appuyant sur les connaissances acquises en cour/sur vos lectures. Pour la rédaction, respectez un plan « à la française ». Schließlich wurden Kriterien zur Unterscheidung der Anforderungen auf den Stufen UNIcert® II und UNIcert® III diskutiert, die unter

dem Titel „Standard-orientierte UNlcert®-Prüfungen“ publiziert sind in: Ó Dúill, Michael; Zahn, Rosemary; Höppner, Kristina D.C. (Hg.): *Zusammenarbeiten*. Festschrift für Bernd Voss. Bochum: AKS-Verlag 2005, S. 43-86.

## Bericht der Arbeitsgruppe 4 „Verknüpfung der Prüfungsteile Leseverstehen und schriftliche Textproduktion (UNlcert® II und III)“

Ziel: Erarbeitung eines Konzeptes zur Verknüpfung der Prüfungsteile Leseverstehen und schriftliche Textproduktion (Erstellung handlungsorientierter Prüfungsmaterialien und Bewertung).

### **I Vorüberlegungen und Diskussion zur Verknüpfung der Prüfungsteile Leseverstehen und schriftliche Textproduktion : von den einzelnen Fertigkeiten zu den integrierten Fertigkeiten**

Ausgehend von einer Bestandsaufnahme der Entwicklung des Fremdsprachenunterrichts (vom kommunikativen zum handlungsorientierten Ansatz) und der verschiedenen Formen der Bewertung - von den bisherigen UNlcert®-Prüfungen zum handlungsorientierten Testen – GULT (Guideline for University- Language Testing)- wurden Vergleiche zwischen den verschiedenen Institutionen und ihrer Praxis gezogen und Erfahrungen ausgetauscht. Die 13 Teilnehmer/Innen klärten zunächst, ob Versuche oder Modelle bereits existieren oder von ihnen praktiziert werden, die die beiden Kompetenzen (Leseverstehen und schriftliche Textproduktion) verbinden. Es zeigte sich, dass meist traditionelle Verknüpfungsformen (Zusammenfassungen, Berichte) benutzt werden und oft die Handschrift der verschiedenen „pädagogischen Kulturlandschaften“ tragen. So wurde beispielsweise die DALF-Prüfung C1 mit der Übungsform „Synthèse de textes“ (siehe unten) und anschließend die Prüfung TOEFL angeführt, die für die schriftliche Textproduktion eine Checkliste für die Prüflinge vorsieht, so dass diese selbst überprüfen können, ob sie die Inhalte korrekt behandelt haben. Die neuere Form einer handlungsorientierten Prüfung, wie sie auch für das Unicertsystem an einigen Einrichtungen erprobt wird, bietet die Möglichkeit, eine Brücke zwischen integrativen und separaten Ansätzen zu bauen, auch hinsichtlich der Benotung der Fertigkeiten.

In kleinen Gruppen wurden die Vor- und Nachteile einer integrativen Vorgehensweise erörtert. Zu den positiven Aspekten hinsichtlich der Kombination der beiden Fertigkeiten zählt, dass die produktiven Fertigkeiten durch die rezeptiven unterstützt werden. Die Textvorgaben helfen dem Lernenden beim Schreiben. Die Texte bringen Argumente und fördern den kritischen Verstand. Andererseits gibt es nicht geringe negative Auswirkungen, wenn ein Verständnisfehler beim Lesen entsteht und somit eine schlechte schriftliche Produktion nach sich zieht. Ein Text oder Texte, die falsch verstanden worden sind, führen zu Fehlinterpretationen, und in einem handlungsorientierten Ansatz kann dadurch das Lösen der Aufgabe scheitern. Es besteht also das Risiko, dass das Thema nicht adäquat behandelt wird.

## **II Analyse von bestehenden Bewertungsrastern und Erstellen eines Rasters für kombinierte Fertigkeiten, das handlungsorientierten Prüfungen gerecht wird**

Welche Prüfungs- und Bewertungsmodelle stehen uns zur Verfügung, welche finden wir geeignet? Was möchten wir entwickeln?

Die erst an wenigen Einrichtungen erfolgte Einführung handlungsorientierter Prüfungsmodelle bei UNIcert® sowie die damit z.T. zusammenhängenden möglichen integrativen Prüfungsformen wurden diskutiert. In diesen Prüfungen sind in der schriftlichen Produktion so genannte Fallstudien eingebaut, eine Sammlung von Ressourcen (Texten und anderen Informationsquellen), die als Stütze für eine Meinungsbildung oder Argumentation fungieren oder zur Problemlösung führen. Die Möglichkeit, Aspekte des Leseverstehens indirekt in die Bewertung der schriftlichen Textproduktion einfließen zu lassen, ist gegeben. Noch ist eine gezielte Bewertung der kombinierten Fertigkeiten Neuland, Vorbilder fehlen weitgehend. Für die Auseinandersetzung wurden Beispiele von handlungsorientierten UNIcert® - Prüfungsmaterialien (Braunschweig, Göttingen, Hohenheim) herangezogen. In diesem zweiten Teil wurden dann die TeilnehmerInnen aufgefordert, gezielt für die Bewertung kombinierter Fertigkeiten (Leseverstehen und schriftliche Textproduktion), ein



Bewertungsraster zu konzipieren. Das Bewertungsraster sollte für den Prüfungsteil Textproduktion einer handlungsorientierter Prüfung erstellt werden. Zur Bearbeitung standen die folgenden Skalen oder Raster zur Auswahl: Die Tabelle der möglichen kombinierten Fertigkeiten aus dem „Manual for relating Language Examinations to the Common European Framework of Reference for Languages (CEFR), (A3.4- Integrated Skills, Form A15: Integrated Skills Combinations, Form A16: Integrated Skills, s.139), die Deskriptoren der Stufen B2 und C1 des Gemeinsamen Europäischen Referenzrahmens für Sprachen, die im Vorfeld aus den verschiedenen Skalen und Subskalen herausgefiltert worden waren, die ALTE (Association of Language Testers in Europe)-Deskriptoren (Individual Component Checklist), die DIALANG-Deskriptoren für das Schreiben und das Lesen, die ELP (European Language Portfolio) - Deskriptoren (A bank of descriptors for self-assessment in European Language Portfolios, 2004) sowie die Bewertungsraster der DELF B2 und der DALF C1 -Prüfungen.

Vor der Bearbeitung erfolgte eine Präzisierung: Das Material war als sehr umfangreiche Dokumentation zu verstehen und somit musste explorativ ausgewählt werden. Die TeilnehmerInnen hatten die Aufgabe, die geeigneten Deskriptoren zu selektieren und sie für kombinierte Fertigkeiten zu bearbeiten.

Die folgenden Punkte wurden dabei teilweise kontrovers diskutiert:

- Die ALTE-Deskriptoren wurden von einem Teil der Gruppe als zu allgemein (Tourismus, Beruf und Studium) befunden, während ein anderer Teil die beiden Bereiche (Beruf und Studium) als sehr relevant betrachtete.
- Es wurde auch festgestellt, dass es für manche Deskriptoren der Subskalen des GERS keine erheblichen Unterschiede zwischen B2 und C1 gibt ⇒ Daraus resultierte die Frage, ob es dementsprechend schwierig sei, differenzierte Bewertungsraster für diese Niveaus zu erstellen?
- Es wurde auch über das Verhältnis zwischen den Fertigkeiten bezüglich der Note diskutiert. Was soll der Prüfer bei kombinierten Fertigkeiten (vor allem) berücksichtigen? Die Lesefähigkeit? Die Schreibfähigkeit? Die grammatische Korrektheit, die adäquate

Verwendung des Wortschatzes, die Vielfalt des Ausdrucks? Oder die Bewältigung der Aufgabe, die (sehr) gut oder weniger gut –z.B. wegen Fehlinterpretationen oder mangelnder Sprachbeherrschung- gelungen sein kann?

- Das Kriterium „Aufgabenbewältigung“ wurde als sehr wichtig betrachtet. Eine längere Auseinandersetzung mit der konkreten Bewertung dieses Kriterium erschien der Gruppe als eine dringende Aufgabe auf dem Weg zur Einführung handlungsorientierter Prüfungen in größerem Maßstab.

- Im Anschluss an die o.g. Diskussionspunkte wurde ein französisches Modell gezeigt. Es handelt sich um das Bewertungsraster der schriftlichen Produktion der DALF-Prüfung C1, die zum Beispiel neben den lexikalischen und grammatikalischen Kriterien weitere Deskriptoren verwendet, die kombinierte Fertigkeiten berücksichtigen: z.B. einen Deskriptor zur inhaltlichen Angemessenheit bei der Übung „Synthèse de textes“ - „Die Darstellung ist objektiv (keine textfremden Aspekte verwendet)“ und einen weiteren zur Textbearbeitungskompetenz – „Kann die aussagekräftigsten Informationen und Argumente aus verschiedenen Quellen zusammenfassen und wiedergeben“. In C2-Niveau gibt es einen Deskriptor zur Fähigkeit, einen Text über ein komplexes Thema zu erfassen: „Kann eine Problematik perspektivisch darstellen, sie durch Haupt- und Nebenargumente untermauern sowie diese durch aussagekräftige Beispiele aus den bereitgestellten Texten und aus persönlichen Erfahrungen erläutern“.

-Es wurde danach deutlich, dass eine eigene, maßgeschneiderte Kombination der Deskriptoren eine adäquate Lösung zu sein scheint wie das obige Beispiel verdeutlicht. In dieser Hinsicht ist es ratsam, die Subskalen der rezeptiven und produktiven Fertigkeiten nach eigenem Nutzen zu kombinieren, denn „fertig kombinierte“ Deskriptoren gibt es in dem Sinne nicht. Als Hilfe für diese Arbeit stehen die Formulare des „Manual for relating Language Examinations to the Common European Framework of Reference for Languages (CEFR)“ zur Verfügung, die für die jeweiligen Fertigkeiten aber auch für kombinierte Fertigkeiten (A3.4- Integrated Skills, Form A15: Integrated Skills Combinations, Form A16: Integrated Skills, s.139) einen Bezug zu den Subskalen

erstellen. Darüber hinaus sollte man in einem handlungsorientierten Ansatz der Dimension der Aufgabenbewältigung besondere Aufmerksamkeit schenken und die Deskriptoren dementsprechend adaptieren.

Da die Fragen und meinungsbildenden Diskussionen zum handlungsorientierten Ansatz bei UNlcert®-Prüfungen als sehr wichtig erachtet wurden und entsprechend viel Zeit beanspruchten, konnte das Bearbeitungsziel bezüglich der Bewertungsraster nicht ganz erreicht werden. Indes lieferten die methodischen Diskussionen viele Hinweise und Anregungen und bereiteten den dritten Teil der AG vor.

### **III Experimentelle Prüfungskonzeption in den jeweiligen Sprachen: Englisch, Französisch, Schwedisch, Deutsch, Spanisch**

Es wurden Gruppen nach Sprachenzugehörigkeit der TeilnehmerInnen (Englisch, Französisch, Schwedisch, Spanisch, Deutsch) gebildet. Diese wurden aufgefordert, eine handlungsorientierte Prüfung nach dem GULT-Modell schematisch aufzubauen. Diese sollte erstens das Hörverstehen und das Leseverstehen durch getrennte Audiotexte und Lesetexte vorsehen und dann die Textproduktion in der Form einer Mini-Fallstudie mit Aufgabenstellungen und Lösungsansätzen beinhalten.

Dank einer Internetrecherche nach Texten und Videos war es möglich, anhand von authentischen Materialien Ideen für die Gestaltung der Prüfung zu bekommen. Die üblichen Selektionskriterien (Länge des Dokuments, verständliche Sprache etc.) wurden bei der Auswahl berücksichtigt. Große Aufmerksamkeit wurde der Suche nach einer adäquaten Problematik für die Fallstudie gewidmet. Während Internet und Zeitungen meist eine Fülle von aktuellen Kontroversen oder Problemen bieten, die sich für das Konzipieren von Fallstudien eignen, haben dennoch die meisten Lehrkräfte zunächst Schwierigkeiten, eine geeignete Aufgabenstellung zu finden und zu formulieren.

Hierbei scheint es angeraten zu sein, sich von den Aufgabenstellungen im klassischen Sinne zu lösen und stattdessen entsprechend aktueller Kommunikationssituationen und

-bedürfnisse ein authentisches Problem zu finden bzw. zu konstruieren, das die Prüflinge möglichst realitätsnah lösen sollen, was anders als bei herkömmlichen Prüfungen zugleich eine zusätzliche Motivation bei der Problemlösung ist und sogar hilft, den Druck der Prüfungssituation etwas abzubauen oder gar in den Hintergrund treten zu lassen.

So könnte etwa statt einer Aufforderung zur Erörterung des Pro und Kontra der E-Books im Sinne einer traditionellen Schreibaufgabe eher die praktische, konkrete Ebene des Buchhändlers oder Nutzers als Ausgangssituation betrachtet werden, die mühelos Ansätze zur Formulierung einer Reihe von möglichen Problemen bietet, die vom Prüfling weitgehend so gelöst werden sollen, wie er es auch im wirklichen Leben tun würde

## **Bilanz**

Als positiv wurde die Auseinandersetzung mit neueren Perspektiven der Handlungsorientierung sowie mit der Bewertung handlungsorientierter Prüfungen betrachtet.

Die TeilnehmerInnen berichteten nachträglich über eine Erweiterung der Kenntnisse über handlungsorientiertes Unterrichts- und Prüfungsmaterial. Es sind sowohl Aspekte der Einsatzmöglichkeiten (z.B. ab wann man mit Fallstudien unterrichten und prüfen kann: ein Teil der Gruppe war der Meinung, dass es im Anfangsunterricht schon möglich ist, während ein anderer Teil die Nutzung mehr für Fachsprachen in den höheren Stufen sah) als auch methodische Aspekte der Konzeption des Materials (Aufgabenstellung, Wichtigkeit der Problemstellung) angesprochen worden. Beispiele zeigten, dass einige Institutionen positive Erfahrungen damit gemacht hatten. Es wurde jedoch auf eine Einschränkung hingewiesen: Die Textlänge bei manchen Prüfungen (z.B. einigen sehr langen Fallstudien) könnte problematisch sein.

Insgesamt wird die integrative Art der handlungsorientierten Prüfung als interessant eingeschätzt, wobei jedoch für die Bewertung noch Erfahrungen gesammelt werden müssen.

Die Frage nach Wegen zu geeigneten Bewertungsrastern für kombinierte, integrierte Fertigkeiten war in dem Workshop von zentraler Bedeutung und hat nicht nur das Problembewusstsein für das Arbeiten mit passenden Deskriptoren und den entsprechenden Bewertungskriterien geschärft, sondern auch Gewohnheiten in Frage gestellt. Sollte man von Texten ausgehen und dann die entsprechenden Deskriptoren festlegen oder wäre es nicht sinnvoller, erst einige Deskriptoren auszuwählen, diese miteinander zu verbinden und erst dann die zur Bewertung der einschlägigen Fertigkeiten geeigneten Themen und Texte aussuchen?

Da das gesamte Workshopmaterial auf der Plattform Moodle hinterlegt worden und weiterhin für die Teilnehmer zugänglich ist, bietet sich hier Gelegenheit für eine Fortsetzung des Austauschs und der Diskussionen. Die TeilnehmerInnen konnten schon mit zusätzlichem Material den Input des Workshops ergänzen. Vielleicht kann auf diesem Wege auch der integrative Prüfungsansatz gemeinsam weiter entwickelt werden.

Catherine Jaeger, Sylvie Richard

## **ARBEITSGRUPPE 5: OFFENE PRÜFUNGSAUFGABEN: ERSTELLEN UND BEWERTEN: EINE EINFÜHRUNG FÜR UNICert®-NEULINGE**

**Claudia Duttlinger, Eleanor Freer - Eberhard-Karls-Universität Tübingen**

Die AG bestand aus Teilnehmern mit verschiedenen sprachlichen Hintergründen und mit unterschiedlichem Wissensstand bzgl. des UNICert®-Konzeptes und seinen Anforderungen: Dementsprechend waren unterschiedliche Themenbereiche von Interesse; z.B.:

- Einführung in das UNICert®-Konzept
- Beschreibung verschiedener Module
- Empfohlene Lehrwerke
- Einstufungstests
- Prüfungs- und Bewertungsmethoden / Musterprüfungen
- Fragen zur (Re-)Akkreditierung
- Fach- und stufenspezifische Fragestellungen und Problemfelder.

Diese Themenauswahl kam den individuellen Bedürfnissen der Teilnehmer/-innen entgegen, da auf diese Weise konkrete Fragen beantwortet und bestimmte Problemfelder erörtert werden konnten. Die jeweilig präsentierten Materialien wurden, sofern erwünscht, nach der Tagung an die AG-Teilnehmer versandt.

Persönlicher Kontakt wurde als wichtig geäußert, zum einen, um auf die o.a. individuellen Fragestellungen einzugehen, zum anderen, um Universitätsspezifika (z.B. bzgl. Positionierung des Sprachzentrums und entsprechende Auswirkungen) zu erörtern. Darüber hinaus ist überlegenswert, ob die UNICert-Homepage im Hinblick auf diese typischen, immer wieder gestellten Fragestellungen ergänzt werden soll.

Günstig für die Vorbereitung einer Akkreditierung ist sicher auch der Kontakt zu und die Hilfestellung einem/s bereits akkreditierten Sprachzentrum(s) im Umkreis.