

# MODUL 4

## ASPEKTE DER BEWERTUNG UND AUSWERTUNG

### 1. EIN FAIRES ERGEBNIS ERREICHEN

### 2. RELIABILITÄT

- a) In objektiv ausgewerteten Tests
  - i) Interne Faktoren
  - ii) Externe Faktoren
- b) In subjektiv bewerteten Tests

### 3. EINIGE ASPEKTE BEI DER BEWERTUNG UND AUSWERTUNG OBJEKTIVER TESTVERFAHREN

- a) Wie können Tests mit objektiven Testverfahren ausgewertet werden?
  - i) Auswertung per Computer
  - ii) Auswertung mit Antwortschablone
- b) Rohwerte in Testergebnisse umrechnen

### 4. EINIGE ASPEKTE DER SUBJEKTIVEN LEISTUNGSBEWERTUNG

- a) Wie kann man die produktiven Fertigkeiten fair bewerten?
  - i) Die Materialien und das Testformat
  - ii) Die Bewertungsverfahren
- b) Tests zum Schreiben
  - i) Bewertungsverfahren
  - ii) Beurteilertraining
  - iii) Mehrfachkorrektur
- c) Tests zum Sprechen
  - i) Worin unterscheidet sich die Bewertung schriftlicher Leistungen von der Bewertung mündlicher Leistungen?
  - ii) Aus welchen Phasen besteht ein Prüfertraining?
  - iii) Aspekte der Fairness bei der Bewertung mündlicher Leistungen

**ANHANG A** – Formeln zur Ermittlung der Reliabilität eines Tests

**ANHANG B** – Beispiel für einen Subtest mit Bewertungsvorgabe

**ANHANG C** – Beispiel für einen Subtest zum Schreiben mit Bewertungsanleitungen und Musterkorrekturen

## 1. Ein faires Ergebnis erreichen

Testautoren könnten der Ansicht sein, dass ihre Arbeit im Testerstellungsprozess beendet ist, wenn die Items begutachtet, erprobt und in die Itembank aufgenommen wurden, um sie in späteren Tests zu verwenden. Was bei einem Test im Echteinsatz geschieht, wie die Ergebnisse beurteilt und die erreichten Punktwerte in Noten umgerechnet werden, um sie den Kandidaten mitteilen zu können, ist ein weites Feld, das für Testautoren möglicherweise nicht weiter von Bedeutung ist.

### Inwiefern ist der Testautor an den Bewertungs- und Auswertungsverfahren eines Tests beteiligt?

Der Testautor ist dafür verantwortlich, gute Items und eindeutige Bewertungsvorgaben zu schreiben. Dies hat einen großen Einfluss darauf, wie leicht ein Item bewertet werden kann und ob Fehler bei der Bewertung (und die damit verbundenen unfairen Ergebnisse) vermieden werden können. Der Testautor hat damit direkt oder indirekt Einfluss auf den *gesamten* Testprozess.

Der Testautor muss die nachfolgenden Phasen der Testdurchführung verstehen, damit Rückmeldungen darüber, wie gut die Items und die Bewertungsvorgaben im Echteinsatz funktionieren, für ihn verständlich sind und ihm zukünftig beim Schreiben von Items helfen. Die Rückmeldungen aus dem Echteinsatz helfen auch der Testentwicklung insgesamt, da sie Aufschluss darüber geben, wie eine bestimmte Fertigkeit getestet werden sollte.

Es ist deshalb wichtig, ein **Rückmeldesystem** zu etablieren. Rückmeldungen können z.B. regelmäßig bei Testredaktionssitzungen besprochen werden oder den Testautorenteams auch in schriftlicher Form gegeben werden. Eventuell ist eine Rückmeldung an das Team angemessener als an individuelle Autoren, da die Items den Begutachtungsprozess mit Kommentaren von mehreren Testautoren durchlaufen haben.

### Was sind die wichtigsten Aspekte bei der Bewertung und Ergebnisermittlung?

Bei jeder Art von Test gehört es zu den Pflichten des Testanbieters, dem Kandidaten ein **faires Ergebnis** mitzuteilen. Dabei geht es in erster Linie darum, Fairness dadurch zu erreichen, dass großer Wert auf die Genauigkeit und Konsistenz der Bewertung gelegt und der Messfehler möglichst gering gehalten wird.

Betont man die Fairness der Bewertung, spielt die **Reliabilität** eine Rolle. Sie ist bei Tests ein besonders wichtiger Aspekt. In diesem Modul wird deshalb eine Reihe von Fragen diskutiert, die mit der Reliabilität (Zuverlässigkeit) der Leistungsmessung zusammenhängen.

Bei der Diskussion der Bewertungsverfahren und Probleme, die damit verbunden sind, können Testaufgaben folgendermaßen klassifiziert werden:

- Aufgaben mit einem Format, das eine objektive Auswertung ermöglicht
- Aufgaben, die wegen ihres Formats eher subjektiv bewertet werden können

In diesem Modul werden verschiedene Fragen diskutiert, die sowohl die objektive Auswertung als auch die subjektive Bewertung betreffen.

## 2. Reliabilität

### a) In objektiv ausgewerteten Tests

Die Reliabilität eines Tests wird oft definiert als die Konsistenz der Leistungsmessung, d. h. eine Messung sollte das, was sie messen möchte, immer gleich messen. Wenn ein Kandidat kurz hintereinander zwei verschiedene Versionen eines Tests ablegt, erwartet man, dass seine Resultate etwa gleich sind; man erwartet nicht, dass eine Version erheblich schwieriger ist als die andere.

#### Welche Faktoren beeinflussen die Reliabilität objektiver Tests?

##### i) Interne Faktoren

Interne Faktoren beziehen sich auf die Anzahl und die Qualität der Items in einem Test. Es gibt verschiedene statistische Verfahren, mit denen die Reliabilität des Tests als innere Konsistenz berechnet wird. Diese Verfahren geben uns Korrelationskoeffizienten an; die bekanntesten sind der Alpha-Koeffizient und Kuder-Richardson 20 (KR-20). Die Formel zur Berechnung dieser Koeffizienten findet sich in Anhang A. Berechnungen dieser Art werden vom Computer mithilfe statistischer Programme wie SPSS oder dem klassischen Analyseprogramm MicroCAT durchgeführt.

##### ii) Externe Faktoren

Externe Faktoren, die die Reliabilität beeinflussen können, sind:

- die äußeren Bedingungen der Testsituation

Das Testergebnis kann durch die Beschaffenheit des Raumes, in dem der Test stattfindet (Licht, bequeme Sitzgelegenheit, Platz, Lärm von draußen), beeinflusst werden. Die Aufsichtsperson kann klare und eindeutige Anweisungen geben oder unsympathisch und abweisend sein. Solche Faktoren sind vermeidbar; man sollte bei der Schulung und Vorbereitung der Testdurchführung in den Testzentren darauf hinweisen.

- Einstellung und Verhalten der Kandidaten

Das Testergebnis könnte auch dadurch beeinflusst werden, wie zuversichtlich bzw. nervös die Kandidaten sind, ob sie sich gesund oder krank fühlen oder ob sie eine starke Abneigung gegen Tests haben; letzteres kann der Grund sein für unkooperatives Verhalten und die Weigerung, in der mündlichen Prüfung ausführlich zu antworten. Die Reliabilität der Leistungsmessung wird auch durch Kandidaten beeinflusst, die bei Multiple-Choice-Items die Lösungen raten, und durch Kandidaten, die entweder keine Erfahrung oder ausgiebige Erfahrungen mit dem Format und den Itemtypen im Test haben.

## b) In subjektiv bewerteten Tests

Die Reliabilität bei subjektiv bewerteten Tests betrifft die Qualität der Bewertungen durch die Beurteiler. Angestrebt werden eine gute

- **Interrater-Übereinstimmung**  
(Übereinstimmung zwischen verschiedenen Beurteilern bei der Bewertung der gleichen Leistung)
- **Intrarater-Übereinstimmung**  
(Übereinstimmung eines Beurteilers mit sich selbst bei der Bewertung unterschiedlicher Leistungen in unterschiedlichen Situationen)

Natürlich ist die subjektive Bewertung stärker von menschlicher Unbeständigkeit und menschlichen Fehlern beeinflusst, als dies bei einer Computerauswertung oder einer mechanischen Auswertung der Fall ist. Es wäre deshalb unrealistisch, hier eine vollkommene Reliabilität zu erwarten. Es gibt jedoch verschiedene Maßnahmen, z. B. ein ausführliches Prüfertraining, mit denen man sicherstellen kann, dass subjektiv bewertete Tests so objektiv und zuverlässig wie möglich sind.

Diese Maßnahmen werden weiter unten in Kapitel 4 über die Bewertung des Schreibens und des Sprechens behandelt.

## 3. Einige Aspekte bei der Bewertung und Auswertung objektiver Testverfahren

### Auswertung der Kandidatenantworten

Das Lese- und Hörverstehen sowie die Sprachstrukturen werden üblicherweise mit itembasierten Tests geprüft. Die verwendeten Itemtypen sind Auswahlitems wie Multiple-Choice, Richtig/Falsch, Lückentests, Zuordnungen und die eher gesteuerten Formate von offenen Aufgaben, wie z. B. Kurzantwortaufgaben.

Für diese Itemtypen gibt es entweder nur eine richtige Antwort, wie z. B. bei Multiple-Choice- und Richtig/Falsch-Items, oder eine sehr begrenzte Zahl möglicher Antworten (die in der Bewertungsvorgabe aufgeführt werden), wie z. B. bei Umformungsaufgaben oder Lückentests mit offenen Lücken.

### a) Wie können Tests mit objektiven Testverfahren ausgewertet werden?

- Tests mit geschlossenen Items können mithilfe eines **Lösungsschlüssels** entweder durch den Computer oder durch Hilfskräfte ausgewertet werden, da bei den Items keine Entscheidungen über die Qualität der Lösung erforderlich sind.
- Tests mit kurzen, vom Kandidaten produzierten Antworten können von Hilfskräften mithilfe von **Antwortschablonen** bewertet werden (im Gegensatz zu Beurteilern/Prüfern bei der Bewertung des Sprechens und des Schreibens).

### i) Auswertung per Computer

Tests mit einem Lösungsschlüssel können vom Computer ausgewertet werden. Dies setzt einen Scanner voraus, der die Antworten der Kandidaten erfasst. Dazu müssen die Kandidaten ihre Antworten auf einen maschinenlesbaren Antwortbogen schreiben. In diesem Zusammenhang werden auch Strichcodes auf die Testunterlagen (mit Informationen zu Testzentrum, Kandidat usw.) gedruckt.

Im Falle von Items mit nur einer Lösung (wie Multiple-Choice-Items), kann der Scanner die Antworten direkt vom ausgefüllten Antwortbogen der Kandidaten einlesen und diese Informationen an den Computer weitergeben, der dann den Gesamtpunktwert berechnet. Es wird auch immer häufiger Software eingesetzt, die einzelne Wörter oder Sätze mit einer sehr begrenzten Anzahl möglicher Lösungen abgleichen kann (z. B. Items zum Informationstransfer oder Lücken, die in einem Formular ausgefüllt werden müssen); der Kandidat muss das Wort (oft in Großbuchstaben) oder die Zahl in das entsprechende Feld im Antwortbogen schreiben.

Der Einsatz von Technologie und Computern bei der Auswertung bedeutet, dass bei solchen Tests das Ergebnis mit einem hohen Grad an Präzision berechnet werden kann.

## ii) Auswertung mit Antwortschablonen

Die Auswertung von Tests mit objektiv zu bewertenden Items kann auch von Hilfskräften durchgeführt werden, sofern die Items nicht zu viele Antwortvarianten zulassen. Die Hilfskräfte benötigen – im Gegensatz zu den Prüfern und Beurteilern – für die Auswertung der Tests keine besondere fachliche Qualifikation oder Erfahrung mit Sprachtests. Sie müssen nur den Bewertungsvorgaben, die alle richtigen Lösungen aufführen sowie den richtigen Punktwert für jedes Item angeben, genau folgen.

### Was ist in diesem Prozess die Rolle des Testautors?

Teil der Aufgabe eines Testautors ist es, voranzuschauen und sich zu überlegen, wie der Test ausgewertet wird. Er sollte keine Items erstellen, die bei der Auswertung mit Sicherheit Probleme bereiten.

In einem Lückentest mit offenen Lücken darf man in den Text z. B. keine Lücken mit mehreren möglichen Lösungen setzen, wie bei folgendem Item:

A wonderful new (1) ..... has just been published.

Bei diesem Item sind die Lösungsmöglichkeiten vielfältig (book, magazine, edition, title, song, picture, report, work, CD-ROM usw.). Das Item könnte zwar als Wortschatzübung im Unterricht nützlich sein, als Item in einem Test hingegen ist es ungeeignet (und wird mit Sicherheit bei der Begutachtung abgelehnt).

Deshalb ist es wichtig, dass der Testautor schon beim Entwerfen der Items klare und korrekte Bewertungsvorgaben (Antwortschablonen) erstellt, um sicherzustellen, dass er nur Items mit einer geringen Anzahl möglicher Lösungen auswählt. Bei dieser Art der Auswertung durch Hilfskräfte müssen Vorgaben zur Verfügung gestellt werden, die alle akzeptablen Lösungen aufführen. Die Hilfskräfte sollen nicht gezwungen sein, selbst zu entscheiden, ob eine Lösung (noch) akzeptabel ist oder nicht.

Das Beispiel unten zeigt einige Items und die dazugehörigen Bewertungsvorgaben:

|   |
|---|
| <p>(a) Please do not smoke in this area of the restaurant.</p> <p>Customers are requested .....</p> |
| <p>(b) Although he took a taxi, Bill still arrived late for the concert.</p> <p>In spite .....</p>  |
| <p>(c) Carol finds it easy to make friends.</p> <p>Carol has no .....</p>                           |

| BEWERTUNGSVORGABE   |
|---|
| <p>(a) not to smoke    ODER    to refrain from smoking</p>  |
| <p>(b) of taking/having taken    a taxi</p> <p>the fact that he/Bill    hired/had hired/took/ had taken    a taxi</p> <p>he/Bill    (still) arrived late.</p> |
| <p>(c) problem(s)/difficulty/trouble (in)    making friends/forming relationships</p>   |

Die Kandidaten müssen wissen, was von ihnen erwartet wird. Bei den Items (b) und (c) oben gibt es mehrere mögliche Lösungen, so dass der Kandidat unsicher werden könnte und dadurch Zeit verliert, dass er sich überlegt, welches nun die *wirklich* richtige Antwort ist.

Eine Möglichkeit, bei Transformationsaufgaben die richtigen Lösungen einzugrenzen, besteht darin, den Anfang und das Ende des Satzes vorzugeben und darauf zu achten, dass die richtigen Lösungen für die Lücke eingeschränkt sind. Im folgenden Beispiel variieren die möglichen Lösungen je nach Lücke zwischen zwei und fünf Wörtern, von denen eines das vorgegebene Wort ist.

|   |
|---|
| <p>(a) Mike's father started the company that Mike now runs.</p> <p>set</p> <p>The company that Mike now runs ..... his father.</p>             |
| <p>(b) I don't recommend hiring skis at this shop.</p> <p>advisable</p> <p>It's ..... skis at this shop.</p>                                    |
| <p>(c) 'This is the best hotel I've ever stayed in,' my colleague said.</p> <p>never</p> <p>I've ..... hotel than this,' my colleague said.</p> |

|  |
|--|
| <b>BEWERTUNGSVORGABE</b>                         |
| (a) was set up by                                |
| (b) not advisable / advisable not to hire (your) |
| (c) never stayed in a better                     |

**Weitere Beispiele:** Ein kompletter Subtest (*Use of English*) aus der englischen Prüfung *First Certificate in English* mit Items und Bewertungsvorgaben findet sich in Anhang B.

### Wichtige Aspekte bei der Auswertung mit Antwortschablonen

- Die Bewertungsvorgaben vervollständigen

Testautoren können nicht immer alle möglichen Antworten voraussehen, die die Kandidaten schreiben werden. Bei Items mit Kurzantworten geschieht es ziemlich häufig, dass die Auswerter Lösungen finden, die ebenso akzeptabel sind, die aber nicht in die Antwortschablone aufgenommen wurden. In diesem Fall muss die Bewertungsvorgabe entsprechend angepasst werden.

Wird der Test nur in einer Institution durchgeführt, ist es einfach, mit dieser Situation umzugehen. Bei einer nationalen/internationalen Prüfung steht den Teams von Hilfskräften oft ein Gruppenleiter vor, der das Fachwissen hat, um in Zweifelsfällen zu entscheiden, welche Lösung noch akzeptabel ist und welche nicht. Die Bewertungsvorgaben können sich im Verlauf der Auswertungen mehrmals

ändern, um weitere mögliche Lösungen zu berücksichtigen.

- Schulung und Kontrolle

Die Hilfskräfte werden darin geschult, die Bewertungsvorgaben angemessen anzuwenden und auf eine genaue Auswertung zu achten. Trotzdem können Fehler vorkommen, besonders wenn sehr viele Kandidaten an der Prüfung teilgenommen haben. Hilfskräfte können die Bewertungsvorgaben oder die Handschrift eines Kandidaten falsch lesen, einen falschen Punktwert aufschreiben oder die Punktwerte für einen Subtest falsch addieren. Solche Fehler müssen in der Schulung angesprochen werden. Außerdem muss die Arbeit der Hilfskräfte überprüft werden, z. B. indem der Gruppenleiter stichprobenartig ihre Auswertungen kontrolliert.

## b) Rohwerte in Testergebnisse umrechnen

Der Gesamtpunktwert für einen itembasierten Test berechnet sich aus allen richtigen Antworten. Dieser Rohwert muss dann in ein Ergebnis umgewandelt werden, das für die Kandidaten und alle Personen, die mit dem Test zu tun haben, aussagekräftig ist, z. B. für Arbeitgeber oder Universitäten, die aufgrund des Ergebnisses Entscheidungen fällen. Ein Punktwert von 27 aus 50, z. B., ist erst dann aussagekräftig, wenn er zu einem bekannten Bewertungssystem in Beziehung gesetzt wird. Bekannte Bewertungssysteme sind „bestanden/nicht bestanden“, „A, B, C“ oder „sehr gut, gut, befriedigend usw.“.

# 4. Einige Aspekte bei der subjektiven Leistungsbewertung

## Bewertung der Kandidatenantworten

Performanztests, d. h. Tests zu den produktiven Fertigkeiten (Sprechen und Schreiben) sind im Allgemeinen nicht itembasiert, obwohl das Schreiben bis zu einem gewissen Grad auch durch gesteuerte Items geprüft werden kann.

Tests zu den produktiven Fertigkeiten stellen andere Anforderungen an die Bewertung, sie wird daher als subjektiver angesehen.

## Wie kann man die produktiven Fertigkeiten fair bewerten?

Die Qualität von Tests für die Fertigkeiten Sprechen und Schreiben wird von folgenden Faktoren beeinflusst:

### i) Die Materialien und das Testformat

Um in subjektiv bewerteten Tests faire Bedingungen für Kandidaten zu schaffen, sollte man sicherstellen, dass der Test **unterschiedliche Aufgaben** enthält. Die Fähigkeiten eines Kandidaten können je nach vorgegebenem Thema oder gewähltem Aufgabentyp sehr unterschiedlich sein. Deshalb sollte jeder Test zum Schreiben oder zum Sprechen aus mehr als nur einer Aufgabe bestehen. Kandidaten kommen aus unterschiedlichen Altersgruppen, haben unterschiedliche Voraussetzungen und unterschiedliche Interessen; deshalb geben ihnen Wahlmöglichkeiten im Test eine Chance, eine Aufgabe zu finden, bei der sie ihr Bestes geben können. Ein Kandidat schreibt vielleicht lieber eine Erzählung als Sachtexte, die bei einem Bericht erwartet werden. Möglicherweise diskutiert er auch lieber über die Klimaauswirkungen auf das Verhalten der Menschen als über die Auswirkungen des Tourismus, je nachdem aus welchem Land er kommt.

**Zu beachten:** Bei Aufgaben mit Wahlmöglichkeiten kann es schwieriger sein, die Leistungen der Kandidaten zu vergleichen, besonders wenn diese unterschiedliche Textsorten produzieren. Eine Möglichkeit wäre, eine Aufgabe zu stellen, die von allen bearbeitet werden muss, und eine weitere, bei der die Kandidaten eine Wahlmöglichkeit haben.

## ii) Die Bewertungsverfahren

Um in Performanztests eine faire Bewertung sicherzustellen, muss auf Folgendes geachtet werden:

- **Einstellung und Training von Prüfern und Beurteilern**

Für die Bewertung von Performanztests werden gute Prüfer und Beurteiler gebraucht. Bei nationalen und internationalen Prüfungen mit hohen Kandidatenzahlen sind dies oft Lehrer, die die Prüfung kennen, weil sie ihre Lerner darauf vorbereiten. Trainingsseminare tragen dazu bei, sie mit den Bewertungsanleitungen oder den Kriterien, mithilfe derer sie die Leistungen der Kandidaten bewerten sollen, stärker vertraut zu machen.

- **Fundierte Standardisierungsverfahren**

Teil des Prüfer- und Beurteilertrainings ist ein Verfahren, das als „Standardisierung“ bezeichnet wird und das zum Ziel hat, bei allen eine einheitliche Bewertung der Kandidatenleistungen zu erreichen.

- **Klare Bewertungskriterien**

Eine Bewertungsskala kann aus Abstufungen in Form von Zahlen, Buchstaben oder Bezeichnungen, wie „gut“, „befriedigend“ usw., bestehen und enthält für jede Stufe eine Definition der entsprechenden Leistung; dies sind die Deskriptoren. Prüfer und Beurteiler müssen die Prinzipien, die der Bewertungsskala zugrunde liegen, gründlich verstehen, da sie mit dieser Bewertungsskala arbeiten müssen.

- **Kontrolle und Evaluierung von Prüfern und Beurteilern**

Die Leistung und das Verhalten der Prüfer und Beurteiler muss auch evaluiert werden. Deshalb ist es notwendig, ein System einzurichten, in dem Leistung und Verhalten der Prüfer und Beurteiler beobachtet, kontrolliert und bewertet werden können. Außerdem muss ihnen eine konstruktive Rückmeldung gegeben werden.

- **Anzahl der Prüfer und Beurteiler bei der Bewertung**

Die Frage, wie viele Prüfer und Beurteiler in der Bewertungsphase der Prüfung mitarbeiten sollten, betrifft sowohl die Bewertung des Schreibens als auch die des Sprechens.

In diesem Modul werden die oben genannten Aspekte im Hinblick auf Tests zum Sprechen und zum Schreiben näher betrachtet.

## a) Tests zum Schreiben

Bevor ein Test zum Schreiben entwickelt wird, muss entschieden werden, ob der Test primär sprachliche Korrektheit (z. B. hinsichtlich Grammatik, Zeichensetzung und Rechtschreibung) prüfen soll oder ob es Ziel des Tests ist, anhand von Aufgaben, die die Realsituationen widerspiegeln, schriftliche Kommunikationsfähigkeit zu prüfen (z. B. einen Beschwerdebrief über einen problematischen Urlaub schreiben). Je nach der Entscheidung, die in dieser Phase gefällt wird, kann der Input für die Testaufgabe entweder nur aus einer Überschrift für einen Aufsatz oder aus mehreren authentischen Vorlagen bestehen, wie z. B. Werbeanzeigen, Briefen, Auszügen aus Zeitungsartikeln.

In der Bewertung der Leistungen stellt sich bei Tests zum Schreiben die Frage, wie man subjektive Faktoren so weit kontrollieren kann, dass eine faire Bewertung erreicht wird. Abhängig vom vorgegebenen Input und den Zielen des Tests können die schriftlichen Arbeiten entweder nach **Korrektheit** (siehe oben), **analytisch** anhand ausführlicher Kriterien oder **holistisch** (Gesamteindruck) bewertet werden.

## i) Bewertungsverfahren

### ▪ Bewertung der Korrektheit

Der Beurteiler kann jeden Fehler in der Grammatik, der Rechtschreibung und der Zeichensetzung unterstreichen und die schriftliche Arbeit danach bewerten, wie viele Fehler sie enthält. Er kann einen maximalen Punktwert festsetzen und je nach Anzahl der Fehler von diesem Maximum Punkte abziehen. Eine andere Möglichkeit besteht darin, proportional maximale Punktwerte für Grammatik, Rechtschreibung usw. zu vergeben und für Fehler Punkte in der entsprechenden Kategorie abzuziehen.

Diese Art von Bewertungsverfahren hat jedoch den Nachteil, dass es risikofreudige Kandidaten bestraft, die mehr schreiben und eine komplexere Sprache verwenden, und diejenigen Kandidaten belohnt, die kürzere und einfachere Texte schreiben. Es ist schwierig, dieses Bewertungsverfahren so anzupassen, dass es auch das Register des Textes und die kommunikative Gesamtwirkung des Geschriebenen berücksichtigt. Grammatikalische Korrektheit und Rechtschreibung können auch in itembasierten Tests geprüft und objektiv ausgewertet werden.

### ▪ Analytische Bewertung

Bei der analytischen Bewertung vergibt der Beurteiler z. B. Punkte auf einer Skala (z. B. von 1 bis 5, wobei 1 eine sehr schwache und 5 eine sehr gute Leistung ist), nachdem er die schriftliche Arbeit getrennt nach verschiedenen Kriterien bewertet hat, wie z. B. Ausdrucksfähigkeit, grammatische Strukturen oder Angemessenheit des Registers. Der Gesamtpunktwert ist die Summe der Punktwerte, die ein Kandidat in den verschiedenen Kriterien erreicht (oder ein Profil der Kandidatenleistung). Üblicherweise werden bei einer analytischen Bewertung sowohl sprachliche Aspekte (Spektrum der sprachlichen Mittel und Korrektheit) als auch inhaltliche Aspekte (Aufgabenerfüllung) bewertet.

### ▪ Holistische Bewertung oder Gesamteindruck

Der Beurteiler vergibt die Punktwerte danach, wie die Gesamtwirkung der schriftlichen Arbeit bewertet wird. Bei diesem Bewertungsverfahren werden die Punktwerte nicht auf verschiedene Kriterien für den schriftlichen Ausdruck verteilt, aber natürlich werden auch hier bestimmte Kriterien berücksichtigt.

Der Beurteiler vergibt nach mehrmaliger schneller Lektüre der schriftlichen Arbeit einen Punktwert, z. B. auf einer Skala von 1 bis 5 oder 1 bis 20. Die Bewertung basiert auf seinem Wissen darüber, was eine angemessene Leistung auf dem Niveau der Prüfung hinsichtlich Aufgabenerfüllung, Wortschatz und Strukturen ausmacht. Eine zweite Bewertung jeder schriftlichen Arbeit kann sinnvoll sein. Beurteiler können sowohl Fehler als auch gelungene Passagen im Text markieren, damit sie schnell entscheiden können, zu welchem der Deskriptoren von 1 bis 5 die schriftliche Arbeit am besten passt.

Es folgen zwei Beispiele (A und B) mit einer Bewertungsanleitung für die holistische Bewertung des Schreibens, die einer Prüfung für Englisch auf der Niveaustufe B2 entnommen wurden.

## A

|       |   |
|-------|---|
| 18-20 | Aufgabe erfolgreich erfüllt mit einem großen Spektrum sprachlicher Mittel und keinen oder nur minimalen Fehlern.                            |
| 16-17 | Kann mehr produzieren als eine Ansammlung einfacher Sätze. Macht nur gelegentlich Fehler. Aufgabe erfolgreich erfüllt.                      |
| 11-15 | Einfache aber korrekte Ausführung der Aufgabe. Gelegentliche Fehler, die aber nicht vom Inhalt der Aussage ablenken; angemessener Ausdruck. |
| 8-10  | Die Botschaft wird vermittelt, aber Fehler sind auffällig; Aufgabe nicht durchgehend erfüllt.   |
| 5-7   | Ungenügende Sprachbeherrschung und häufige grundlegende Fehler; Aufgabe nur teilweise erfüllt bzw. Anweisung nicht umgesetzt.               |
| 0-4   | Nur bruchstückhafte Sprachäußerungen; Inhalt irrelevant oder nicht zu bewerten, da zu knapp.  |

In der Bewertungsanleitung A wird nicht versucht zu definieren, was z. B. ein Testwert von 12/20 oder 6/20 bedeutet. Es ist schwierig, auf einer Skala mit 20 Punkten solche feinen Unterschiede zu verdeutlichen. Ob Bewertungen mit dieser Skala konsistent sind oder nicht, wird daher sehr vom Training und dem Können der Beurteiler abhängen.

## B

|   |   |
|---|---|
| 5 | <p>Aufgabe voll erfüllt.</p> <ul style="list-style-type: none"> <li>▪ Alle Inhaltspunkte werden in angemessenem Umfang behandelt.</li> <li>▪ Großes Spektrum der Strukturen und des Wortschatzes gemäß der Aufgabe.</li> <li>▪ Minimale Fehler, vielleicht weil der Kandidat/die Kandidatin risikofreudig oder ehrgeizig ist; gut entwickelte Beherrschung der Sprache.</li> <li>▪ Effektiver Aufbau der Gedanken mit einer Vielfalt verbindender Elemente.</li> <li>▪ Register und Format durchgehend der Schreibabsicht und dem Leser angemessen.</li> </ul> <p>Erzielt voll die gewünschte Wirkung auf den Leser.</p>  |
| 4 | <p>Gute Erfüllung der Aufgabe.</p> <ul style="list-style-type: none"> <li>▪ Alle wichtigen Inhaltspunkte werden behandelt; ein oder zwei kleinere Punkte werden möglicherweise nicht aufgegriffen.</li> <li>▪ Völlig angemessenes Spektrum der Strukturen und des Wortschatzes gemäß der Aufgabe.</li> <li>▪ Im Allgemeinen korrekt; Fehler hauptsächlich dort, wo versucht wird, komplexere Sprache zu verwenden.</li> <li>▪ Klarer Aufbau der Gedanken mit angemessenen Verbindungen.</li> <li>▪ Register und Format im Allgemeinen der Schreibabsicht und dem Leser angemessen.</li> </ul> <p>Erzielt die gewünschte Wirkung auf den Leser.</p>  |
| 3 | <p>Ausreichende Erfüllung der Aufgabe.</p> <ul style="list-style-type: none"> <li>▪ Alle wichtigen Inhaltspunkte werden behandelt; einige kleinere Auslassungen.</li> <li>▪ Angemessenes Spektrum der Strukturen und des Wortschatzes, die die Anforderungen der Aufgabe erfüllen.</li> <li>▪ Eine Reihe von Fehlern ist möglich, sie behindern jedoch nicht die Kommunikation.</li> <li>▪ Angemessener Aufbau der Gedanken mit einfachen Verbindungen.</li> <li>▪ Ausreichender, wenn auch nicht immer erfolgreicher Versuch, ein der Schreibabsicht und dem Leser angemessenes Register und Format zu verwenden.</li> </ul> <p>Erzielt im Großen und Ganzen die gewünschte Wirkung auf den Leser.</p>       |
| 2 | <p>Versuch, die Aufgabe zu erfüllen, der aber nicht angemessen gelingt.</p> <ul style="list-style-type: none"> <li>▪ Einige wichtige Inhaltspunkte werden unzureichend oder gar nicht behandelt und/oder es gibt irrelevante Äußerungen.</li> <li>▪ Begrenztes Spektrum der Strukturen und des Wortschatzes.</li> <li>▪ Eine Reihe von Fehlern, die den Leser ablenken und zeitweise die Kommunikation behindern können.</li> <li>▪ Kein angemessener Aufbau der Gedanken; nur selten Verwendung von Verbindungselementen.</li> <li>▪ Kein erfolgreicher/konsistenter Versuch, ein angemessenes Register und Format zu verwenden.</li> </ul> <p>Die Aussage wird dem Leser nicht klar genug kommuniziert.</p> |
| 1 | <p>Schwacher Versuch, die Aufgabe zu erfüllen.</p> <ul style="list-style-type: none"> <li>▪ Erhebliche inhaltliche Auslassungen und/oder häufig irrelevante Inhalte, möglicherweise wegen Nicht-Verstehen der Aufgabe bzw. der Anweisung.</li> <li>▪ Geringes Spektrum des Wortschatzes und der Strukturen.</li> <li>▪ Häufige Fehler, die die Kommunikation behindern; kaum entwickelte Beherrschung der Sprache.</li> <li>▪ Kein Aufbau der Gedanken, keine Verbindungselemente.</li> <li>▪ Kein oder nur geringes Bewusstsein von Register und Format.</li> </ul> <p>Die Äußerung hat eine sehr negative Wirkung auf den Leser.</p>  |

Die Bewertungsanleitung B enthält klarere Definitionen der Kriterien als die Bewertungsanleitung A und erlaubt den Beurteilern, sowohl die verschiedenen sprachlichen Aspekte als auch die kommunikative Gesamtwirkung der schriftlichen Äußerung zu bewerten. Mit dieser Bewertungsanleitung können auch Kandidaten angemessen bewertet werden, die zwar Fehler machen, aber mutig versuchen, auch komplexere Sprache zu verwenden oder die Aufgabe auf kulturspezifische und/oder kreative Art und Weise zu lösen.

## Erfüllung der Aufgabe

Beurteiler, die holistisch bewerten, können sich bei ihren Bewertungen auf allgemeine Bewertungsanleitungen, wie die oben gezeigten, stützen, die sich nicht auf eine bestimmte Aufgabe beziehen und für jede Note eine detaillierte Beschreibung der erwarteten Leistung liefern. Manchmal werden auch spezifische Bewertungsanleitungen für einzelne Testaufgaben entwickelt, die eine detaillierte Beschreibung der Kriterien für Aufgabenerfüllung, Register, Strukturen, Wortschatz usw. enthalten. Ziel all dieser Bewertungsanleitungen ist es, unter den Beurteilern ein einheitliches Urteil darüber zu erzielen, wie gut eine Aufgabe erfüllt wurde.

## ii) Beurteilertraining

Im Allgemeinen kann man sagen, dass die Beurteiler umso qualifizierter und erfahrener sein müssen, je weniger analytisch sie bewerten. Die Beurteiler müssen das Niveau und den Inhalt der Prüfung kennen; dies ist z. B. der Fall, wenn sie ihre Lerner auf die Prüfung vorbereitet haben.

### Aus welchen Phasen besteht ein Beurteilertraining?

Bei einer internen Schulprüfung können die unten genannten Verfahren verkürzt werden, aber auch hier sind die grundlegenden Schritte die gleichen, unabhängig von der Anzahl der Kandidaten.

#### ▪ Standardermittlung

Nachdem der Test durchgeführt wurde, sieht sich der für die Prüfungsdurchführung Verantwortliche\* so viele schriftliche Arbeiten an, wie in der vorgegebenen Zeit möglich ist. Es geht darum, einen allgemeinen Eindruck davon zu bekommen, was geschrieben wurde und ob die Aufgaben im Test zum Schreiben in Ordnung waren oder Probleme bereitet haben. Nach der Auswahl von Beispielen, die für jede Stufe der Bewertungsanleitung repräsentativ sind, trifft er sich mit erfahrenen Beurteilern, um die Arbeiten zu bewerten und die Bewertungen zu vergleichen und zu diskutieren. Dabei wird auch erörtert, wie das Kriterium „Erfüllung der Aufgabe“ bewertet werden sollte. Sobald eine Einigung bei der Bewertung erzielt wurde, wird das Ergebnis begründet und schriftlich festgehalten. Diese bewerteten schriftlichen Arbeiten können dann als Musterkorrekturen (Standard) im Beurteilertraining eingesetzt werden.

#### ▪ Standardisierungsverfahren

In den Trainingsseminaren erhalten die Beurteiler die oben erwähnten ausgewählten schriftlichen Arbeiten (ohne die Bewertungen), um sie zu bewerten und mit dem Teamleiter über ihre Bewertungen zu diskutieren. Diese Phase kann auch außerhalb eines Seminars stattfinden, wenn die Beurteiler zu Hause bewerten und dem Teamleiter ihre Ergebnisse per Telefon mitteilen. Wenn die Beurteiler bei ihren Bewertungen zu streng oder zu milde waren, müssen sie sich dem für die Prüfung festgesetzten Standard annähern. Die in den Musterkorrekturen festgehaltenen schriftlichen Begründungen für die Bewertung können von den Beurteilern falls nötig in der gesamten Phase der Bewertung konsultiert werden. Es ist jedoch wichtig, dass die Beurteiler vor jedem Einsatz als Beurteiler erneut ein solches Training durchlaufen.

Im **Anhang C** finden Sie einen Subtest zum Schreiben mit Musterkorrekturen, die beispielhaft zeigen, wie die Leistungen bewertet und die Beurteiler bei ihren Bewertungen unterstützt werden.

**Zu beachten:** Wenn Beurteiler zum ersten Mal bewerten, wird ihnen ein Einführungsgespräch helfen, um sie mit den technischen Einzelheiten der Korrektur vertraut zu machen (wie Fehler markiert werden, wie schnell bewertet werden soll, wann die bewerteten Arbeiten zurückgeschickt werden müssen, welche Hilfen sie von erfahrenen Prüfern bekommen können usw.).

---

\*Anmerkung der Übersetzerin:

Das Fallbeispiel bezieht sich auf die Prüfungen von Cambridge ESOL. Bei diesem Prüfungssystem gibt es weltweit für jede Prüfung *Chief Examiners*, die für die Durchführungen einer bestimmten Prüfung in ihrem jeweiligen Land verantwortlich sind. Die Prüfer und Beurteiler sind in Teams organisiert, die von einem Gruppenleiter betreut werden, der dem *Chief Examiner* unterstellt ist.

- Kontrolle der Beurteiler

Während der Bewertungsphase können die Beurteiler von einem Gruppenleiter oder einem erfahrenen Beurteiler kontrolliert werden, um festzustellen, ob sie die Leistungen angemessen bewerten. Es ist hilfreich, wenn die Beurteiler in der Anfangsphase der Bewertungen „gelenkt“ werden, damit sie das Niveau der sprachlichen Leistungen anhand der Bewertungsanleitung richtig erfassen. Häufige Eingriffe in die Bewertungen sind allerdings zu vermeiden, da dies die Beurteiler dazu verleiten könnte, ihre Bewertungen ständig zu ändern; sie können dann nicht mehr konsistent bewerten.

- Auswertung des Beurteilerverhaltens

Wenn die Bewertungen abgeschlossen sind, kann mithilfe statistischer Verfahren festgestellt werden, wie konsistent die Beurteiler bei ihren Bewertungen waren. Es ist nützlich, den Beurteilern derartige Rückmeldungen zu geben, damit sie sich bei zukünftigen Bewertungen daran orientieren können.

Nach Abschluss der Bewertungen schreibt der für die Prüfungsdurchführung Verantwortliche einen Bericht über den eingesetzten Test zum Schreiben und die Qualität der Aufgaben. Es ist wichtig, dass diese Art von Rückmeldung an die Testautoren weitergeleitet wird, um die Qualität zukünftiger Entwürfe zu verbessern.

### iii) Mehrfachkorrektur

Eine Mehrfachkorrektur erhöht die Reliabilität. Bei einer holistischen Bewertung der schriftlichen Arbeiten ist es üblich, die Arbeiten von einem zweiten Beurteiler bewerten zu lassen, so dass alle Arbeiten von mehr als einem Beurteiler bewertet werden. Wenn zwei Beurteiler zu sehr unterschiedlichen Bewertungen kommen, wird die Arbeit von einem dritten Beurteiler bewertet. Dies kann der Gruppenleiter oder der *Chief Examiner* sein.

Die Bewertungen können auch in einer gemeinsamen Sitzung mit allen Beurteilern stattfinden, um alle schriftlichen Arbeiten aus einer Prüfungsdurchführung zu bewerten. Dies macht es möglich, das Tempo und die Genauigkeit der Bewertungen zu kontrollieren und dabei statistische Informationen über die Interrater- und die Intrarater-Reliabilität zu erhalten. So können Beurteiler, die inkonsistent bewerten oder die durchgehend strenger bzw. milder bewerten als der Rest der Gruppe, identifiziert werden. Dadurch lässt sich auch feststellen, welche Arbeiten ein weiteres Mal bewertet werden sollten. Darüber hinaus wird deutlich, welche Beurteiler ein weiteres Training brauchen oder welche man in Zukunft nicht mehr beauftragen sollte.

## b) Tests zum Sprechen

### i) Worin unterscheidet sich die Bewertung schriftlicher Leistungen von der Bewertung mündlicher Leistungen?

Vieles, was zur Bewertung des Schreibens gesagt wurde, trifft auch auf die Bewertung des Sprechens zu. Dennoch gibt es einige Unterschiede, die im Folgenden dargestellt werden:

- Die Interaktion mit den Kandidaten

Die Persönlichkeit, das Verhalten und die allgemeine Fähigkeit des Prüfers erhalten bei Tests zum Sprechen, die als Gespräch mit einem Prüfer durchgeführt werden, eine große Bedeutung. Dies ist die einzige Art von Sprachtest, bei dem das Prüfen und das Bewerten simultan stattfinden, außer es werden Ton- oder Videoaufnahmen eingesetzt.

Häufig wird vom Prüfer verlangt, dass er innerhalb von etwa 10 Minuten pro Kandidat gleichzeitig verschiedene Operationen ausführt: Er muss eine Beziehung zum Kandidaten herstellen (während er dabei auch den Input kontrolliert, den er dem Kandidaten gibt), den Kandidaten durch die verschiedenen Aufgaben führen und ihn mit Bezug auf die Bewertungsanleitung bewerten.

- Die zeitliche Planung und das Format des Trainings

Das Training für die Prüfer muss vor der mündlichen Prüfung stattfinden. Vor Beginn der Prüfung müssen die Prüfer mit den Bewertungsverfahren, den Prüfungsunterlagen und den Skalen der Bewertungsanleitung vertraut sein. Es ist sehr wichtig, kompetente Prüfer auszuwählen und zu schulen. Organisationen, die Prüfungen erstellen und durchführen, sollten die Qualifikationen eines Prüfers schriftlich festhalten.

In Trainingsseminaren für Prüfer werden häufig Videoaufnahmen von echten oder simulierten Prüfungen gezeigt, in denen Kandidaten von erfahrenen Prüfern geprüft werden. Die zukünftigen Prüfer sehen sich die Videos an und bewerten die Leistungen. Anschließend werden die Bewertungen mit dem Ziel diskutiert, sich auf eine einheitliche Bewertung zu einigen.

## ii) Aus welchen Phasen besteht ein Prüfertraining?

Das beschriebene Prüfertraining bezieht sich auf das folgende Beispiel eines Tests zum Sprechen auf einer fortgeschrittenen Stufe (C1):

Teil 1: Prüfungsgespräch (5 Minuten)

Teil 2: Paarprüfung. Jeder Kandidat spricht einzeln und stellt oder beantwortet Fragen zu seiner Präsentation. Anschließend nimmt er an einer Diskussion teil (15 Minuten).

- Informationen vor den Trainingsseminaren

Prüfer, die an einem Trainingsseminar teilnehmen, erhalten vor dem Seminar ein Informationspaket, das sie sorgfältig lesen müssen. Die Materialien geben detailliert Auskunft über das Format der mündlichen Prüfung, die Ziele der einzelnen Prüfungsteile, die Durchführungsbestimmungen und darüber, welche Hinweise den Kandidaten gegeben werden sollen usw.

- Trainingsseminare

Die auszubildenden Prüfer nehmen dann z. B. an zwei Seminaren teil, die jeweils einen halben Tag dauern.

In der ersten Sitzung wird der erste Teil der Prüfung, das Prüfungsgespräch, besprochen:

- das Format und das Prüferverfahren werden diskutiert
- Arbeitsblätter werden ausgefüllt, um sicherzustellen, dass die Seminarteilnehmer das System verstanden haben
- die Bewertungsskalen und die Deskriptoren werden vorgestellt
- sechs Videoaufnahmen werden gezeigt, um die Bewertung mündlicher Leistungen zu üben
- die Bewertungen und eventuell aufgetretene Probleme werden diskutiert; anhand von Musterbewertungen wird verdeutlicht, wie die Testwerte für die Kandidaten zustande kamen

In der **zweiten Sitzung** werden verwaltungstechnische Fragen und der zweite Teil der Prüfung besprochen:

- drei Paare mit Kandidaten werden gezeigt, die sich freiwillig für die Video-Aufnahmen zur Verfügung gestellt haben; die Seminarteilnehmer bewerten die Leistungen.
- die Bewertungen und eventuell aufgetretene Probleme werden diskutiert; anhand von Musterbewertungen wird verdeutlicht, wie die Testwerte für die Kandidaten zustande kamen

Manchmal nehmen auch freiwillige Kandidaten an einem Trainingsseminar teil. Sie können dann von den Seminarteilnehmern „geprüft“ werden. Durch diese Übung verstehen die Prüfer und die Kandidaten den Prüfungsablauf besser, inklusive solcher Aspekte wie Zeitmanagement, Sitzordnung, Einsatz der Prüfungsunterlagen, Eintragen der Bewertungen in das Bewertungsförmular usw.

### ▪ Standardisierung

Nachdem die Prüfer geschult wurden, müssen sie regelmäßig an Standardisierungsverfahren teilnehmen, um sicherzustellen, dass ihre Bewertungen konsistent bleiben. Das Standardisierungsverfahren kann aus folgenden Schritten bestehen:

- Für die Standardisierung werden Videoaufnahmen von mündlichen Prüfungen produziert; eine Gruppe erfahrener Prüfer bewertet die Lerner und hält die Begründungen für die vergebenen Bewertungen schriftlich fest.
- Die übrigen Prüfer sehen sich die Videos an und bewerten die Kandidaten anhand schriftlicher Vorgaben.
- Die Prüfer diskutieren ihre Bewertungen, vergleichen sie mit den Bewertungen und Kommentaren dazu auf dem Standardisierungsvideo und passen – wo erforderlich – ihre Bewertungen an.
- Die Prüfer sehen sich eine letzte Aufnahme an. Der Gruppenleiter sammelt danach die Bewertungen ein und prüft sie. Er gibt allen Prüfern, deren Bewertung nicht der Leistungsstufe entsprach, eine entsprechende Rückmeldung.

### iii) Aspekte der Fairness bei der Bewertung mündlicher Leistungen

Es ist sehr wichtig, dass die Prüfer kompetent sind und gut ausgebildet werden. Darüber hinaus gibt es auch andere Faktoren, die die Fairness der Bewertung beeinflussen:

#### ▪ Das Prüferskript

Die Anweisungen im Prüferskript (engl. *interlocutor frame*) steuern die Sprache der Prüfer. Sie sind eine weitere Maßnahme, um den Prüfungsablauf zu standardisieren und ihn so einheitlicher, weniger subjektiv und fairer gegenüber den Kandidaten zu gestalten. Die Videoaufnahmen, die zu Trainingszwecken produziert werden, sollten einen effektiven Umgang mit dem Prüferskript zeigen, z. B. dadurch, dass der Prüfer die Prüfungsfragen so gut kennt, dass er sie nicht ständig von seinem Skript ablesen muss.

Wenn kein Prüferskript eingesetzt wird, müssen die Prüfer darin geschult werden, sorgfältig auf ihre Sprache zu achten und immer daran zu denken, dass sie die Kandidaten zur bestmöglichen Leistung anspornen sollen. Prüfungsgespräche zu üben, die auf Video (oder Tonband) aufgenommen und in Trainingsseminaren diskutiert werden, helfen den Prüfern, Dinge in ihrem Verhalten zu erkennen, die verbessert werden müssen, z. B. der Gebrauch einer Sprache, die über dem Niveau der Prüfung liegt, das Stellen von Fragen, auf die nur mit Ja oder Nein geantwortet werden kann oder ein zu großer Redeanteil während der Prüfung.

Es wird manchmal behauptet, dass ein Prüferskript eine „echte Kommunikation“ zwischen Prüfer und Kandidat verhindert, da der Prüfer nicht spontan auf das antworten kann, was der Kandidat sagt. Diese Kritik muss man abwägen gegenüber der Frage der Fairness, besonders bei internationalen Prüfungen mit großen Kandidatenzahlen, wo die Prüfung in jedem Land hinsichtlich Inhalt und Durchführung gleich sein soll.

#### ▪ Die Bewertungsanleitung

Eine Bewertungsanleitung muss so detailliert sein, dass sie unterschiedliche Leistungen erfassen kann, aber nicht so detailliert, dass sie nicht mehr handhabbar ist in einer Situation, in der sehr schnelle Entscheidungen gefällt werden müssen. Die Leistungsbeschreibungen müssen auf den verschiedenen Stufen klar und eindeutig formuliert sein und sich auf Aspekte der Performanz beziehen, die der Prüfer leicht beobachten kann.

Prüfer müssen sich auch über den 'Halo-Effekt' im Klaren sein. Dieser bedingt, dass gute Leistungen in einem Bereich, z. B. in der Aussprache, den Prüfer dazu verleiten könnten, andere, schwächere Bereiche zu milde zu bewerten. Sie müssen auch vermeiden, einen Kandidaten mit den vorherigen Kandidaten zu vergleichen, und darauf achten, seine Leistung nur im Hinblick auf die Bewertungskriterien zu bewerten.

Das folgende Beispiel zeigt ein Bewertungsraster. Es werden für sechs Kriterien Punkte vergeben, die zusammen das Profil der mündlichen Leistung eines Kandidaten ergeben. Bei jedem Kriterium können 0 bis 5 Punkte vergeben werden, die von Natürlichkeit und Angemessenheit (5) bis zu Unverständlichkeit (0) reichen. Innerhalb eines jeden Feldes gibt es keine Bestehensgrenze für dieses Kriterium, da die Punkte für alle Kriterien zusammengezählt werden und einen Gesamtpunktwert ergeben (aus einem Maximum von 30).

|  |   |   |
|--|---|---|
| <p><b>1. FLÜSSIGKEIT</b></p> <p>Tempo und Rhythmus,<br/>Wahl der Strukturen,<br/>allgemeine Natürlichkeit<br/>und Klarheit</p>   | <p><b>3. AUSSPRACHE</b></p> <p>(Sätze)</p> <p>Betonung, Intonation,<br/>Verbindung von Satzteilen</p>   | <p><b>5. INTERAKTIVE<br/>KOMMUNIKATION</b></p> <p>Flexibilität und Spektrum<br/>der sprachlichen Mittel<br/>beim Austausch von<br/>Informationen und bei der<br/>sozialen Interaktion</p> |
| <p><b>2. KORREKTHEIT<br/>DER GRAMMATIK</b></p> <p>Beherrschung der<br/>Strukturen, inklusive<br/>Tempora, Präpositionen<br/>usw., um wirkungsvoll zu<br/>kommunizieren</p> | <p><b>4. AUSSPRACHE</b></p> <p>(Einzellaute)</p> <p>Korrektter Gebrauch der<br/>Konsonanten und Vokale<br/>in betonter und unbetonter<br/>Position, um das Verstehen<br/>zu erleichtern</p> | <p><b>6. SPEKTRUM DES<br/>WORTSCHATZES</b></p> <p>Abwechslungsreicher und<br/>korrekter Wortschatz im<br/>Kontext der<br/>Kommunikation</p>   |

Das Bewertungsraster oben wird in einer mündlichen Prüfung für Englisch eingesetzt. Die Bewertungsskala, die diesem Bewertungsraster zugrunde liegt, findet sich auf der nächsten Seite.

| Note | Grammatik   | Wortschatz  | Aussprache   | Aufbau   | Kommunikationsstrategien und Interaktion  | Hilfe vom Gesprächspartner | Aufgabenerfüllung   |
|------|---|---|--|--|---|----------------------------|---|
| A    | Großes Spektrum der Strukturen inklusive komplexer Strukturen, die korrekt verwendet werden; sehr wenige Fehler.                                | Sehr großes Spektrum; korrekter und angemessener Gebrauch; Breite der Themen; kaum Zögern bei der Wortwahl.   | Durchgehend korrekter Gebrauch aller Aspekte der Aussprache.   | Ausgezeichnet – logisch entwickelte Äußerungen; präziser Gebrauch von verbindenden Elementen.  | Beabsichtigte Bedeutung wird in allen Kontexten kommuniziert; beginnt Interaktion und hält sie aufrecht; geht mit Leichtigkeit auf Themenwechsel ein.   | Nicht notwendig            | Aufgabe voll erfüllt und erfolgreich kommuniziert.  |
| B    | Vollständiges Spektrum an Grundstrukturen, die mit wenigen Fehlern verwendet werden; Fehler bei komplexen Strukturen.                           | Großes Spektrum des Wortschatzes für die alltägliche Kommunikation, aber nicht immer ausreichend für eine Diskussion; kaum Zögern bei der Wortwahl. | Weitgehend korrekter und durchgehender Gebrauch der meisten Aspekte der Aussprache; muttersprachlicher Einfluss nicht störend. | Gut aufgebaut – die Hauptpunkte werden voneinander abgesetzt und in angemessener Abfolge vorgebracht; meist guter Gebrauch von verbindenden Elementen. | Beabsichtigte Bedeutung wird in fast allen Kontexten kommuniziert; beginnt Interaktion und hält sie aufrecht; gelegentliche Schwierigkeiten, auf Themenwechsel einzugehen.                                      | Gelegentlich notwendig     | Aufgabe erfüllt, aber auf ein oder zwei Anforderungen der Aufgabe wurde nicht eingegangen.    |
| C    | Korrekt Gebrauch der Grundstrukturen; fehlerhafter Gebrauch komplexer Strukturen, der gelegentlich zu Missverständnissen führt.                 | Ausreichender Wortschatz für alltägliche Situationen; geringes Zögern bei der Wortwahl.   | Gelegentlich falsche Aussprache, aber immer noch verständlich; deutlicher muttersprachlicher Einfluss, aber verständlich.      | Eingeschränkte Wirkung; einige verbindende Elemente werden nicht angemessen eingesetzt; gelegentlich Klärungsfragen oder Wiederholungen notwendig.     | Wichtige Gedanken werden kommuniziert; setzt Kompensationsstrategien ein; einige Schwierigkeiten, die Interaktion zu beginnen und auf Themenwechsel einzugehen.   | Häufig notwendig           | Aufgabe nur teilweise erfüllt; auf mehrere Anforderungen der Aufgabe wurde nicht eingegangen. |
| D    | Fehlerhafter Gebrauch vieler Grundstrukturen; seltener und fehlerhafter Gebrauch komplexer Strukturen; häufig Probleme, sich klar auszudrücken. | Eingeschränktes Spektrum des Wortschatzes; nur ausreichend für einfache alltägliche Situationen; häufiges Zögern bei der Wortwahl.                  | Fehler in der Aussprache führen manchmal zu unverständlichen Äußerungen; häufige schwere Fehler behindern die Kommunikation.   | Schlecht aufgebaut – keine verbindenden Elemente; häufig unangemessene Abfolge; nur eingeschränkte Beherrschung der Diskurs- und Gesprächsroutinen.    | Die Hauptgedanken werden nur in eingeschränkten Kontexten kommuniziert; Kompensationsstrategien werden selten eingesetzt; beginnt selten die Interaktion und hat Schwierigkeiten, auf Themenwechsel einzugehen. | Durchgehend notwendig      | Aufgabe nicht erfüllt; es wurde nicht auf die Anforderungen der Aufgabe eingegangen.          |
| E    | Fehlerhafter Gebrauch fast aller Grundstrukturen; keine komplexen Strukturen; weitgehend unverständlich.  | Nur einige Wörter oder kurze Sätze; meistens nicht ausreichend für eine Kommunikation.  | Fehlerhaft und uneinheitlich; weitgehend unverständlich.   | Keine Beherrschung der Regeln des Diskurses oder des Gesprächs; man kann dem Gesagten nicht folgen.  | Große Schwierigkeiten bei der Kommunikation; beherrscht keine Kompensationsstrategien; braucht Partner, um Interaktion aufrechtzuerhalten.  |                            |   |

### ▪ Die Testmaterialien

Dem Prüfer muss genügend Testmaterial zur Verfügung gestellt werden, damit er zu Beginn der Prüfung schnell entscheiden kann, welches das beste Material für den jeweiligen Kandidaten ist. Noch ein weiterer Faktor trägt bei Tests zum Sprechen zu einer fairen Bewertung bei. Wenn z. B. ein Kandidat zu Beginn der Prüfung äußert, dass er sich nicht für Sport interessiert, kann der Prüfer entscheiden, die Materialien zum Thema Sport in den weiteren Phasen der Prüfung nicht einzusetzen.

Bei Tests zum Sprechen müssen die Materialien flexibel eingesetzt werden und die Aufgaben für Kandidaten unterschiedlichen Alters und mit unterschiedlichen Voraussetzungen geeignet sein. Diese Notwendigkeit muss auch den Testautoren bewusst sein.

### ▪ Die Anzahl der Prüfer und ihre Rolle in der Prüfung

Auch die Anzahl der Prüfer, die die Leistungen bewerten, kann die Fairness einer Bewertung beeinflussen.

- Prüfungsgespräche können mit einem Prüfer durchgeführt werden, wobei dieser gleichzeitig das Gespräch leitet und den Kandidaten bewertet.
- Ein Kandidat kann auch von zwei Prüfern geprüft werden, wobei jeder Prüfer die Leistung unabhängig vom anderen Prüfer bewertet.
- Zwei oder drei Kandidaten können von zwei Prüfern geprüft werden, wobei die Prüfer jeweils entweder die Rolle des Gesprächspartners oder die des Beurteilers übernehmen.

Wenn Prüfen und Bewerten voneinander getrennt werden, kann der Beurteiler sich auf die Vergabe einer Bewertung konzentrieren und dabei analytisch bewerten, während der Prüfer als Gesprächspartner das Gespräch leitet und eventuell für jeden Kandidaten eine Bewertung in Form eines Gesamteindrucks vergeben kann. Auf diese Weise erreicht man zwei voneinander unabhängige Bewertungen.

Wenn es möglich ist, dass sich zwei Prüfer an der Bewertung beteiligen, hat dies wegen der größeren Objektivität eindeutige Vorteile gegenüber dem Einsatz eines einzelnen Prüfers; außerdem wird es dadurch einfacher, die Reliabilität sicherzustellen.

Wenn Prüfungsgespräche auf Tonband aufgenommen werden, können mehrere Prüfer zumindest die Kandidaten bewerten, über deren Ergebnis kein Konsens bestand.

### ▪ Die Anzahl der Kandidaten und ihre Beziehung zueinander

Kandidaten können einzeln, in Paaren oder in Gruppen geprüft werden. Die Frage, ob Paare oder Gruppen mit Kandidaten gebildet werden sollen, die einander kennen, oder mit Kandidaten, die einander nicht kennen, ist ein weiterer Aspekt, der die Fairness betrifft. Kandidaten, die einander kennen, sind vielleicht weniger nervös, haben aber eventuell ihre Antworten schon so sehr „geübt“, dass ihre Äußerungen an Natürlichkeit verlieren. Kandidaten, die einander nicht kennen, fühlen sich vielleicht in der Interaktion zunächst verunsichert, sind aber eventuell insgesamt kommunikativer, weil sie ihrem Gesprächspartner besser zuhören müssen.

Probleme mit der Fairness entstehen auch durch unterschiedliche Persönlichkeiten und ihre Wirkung auf die Gesprächspartner. So kann z. B. ein dominierender und gesprächiger Kandidat die Gesprächszeit eines anderen Kandidaten verkürzen. Diese Probleme müssen im Prüfertraining angesprochen werden. Außerdem müssen die Kriterien, die für den Test zum Sprechen gelten, öffentlich bekannt und für Lehrer und Kandidaten leicht zugänglich sein.

## ANHANG A

### Formeln zur Ermittlung der Reliabilität eines Tests

Formel für:

Cronbachs Alpha (Alpha-Koeffizient)

$$\alpha = \frac{k}{k-1} \left[ 1 - \frac{\sum s_i^2}{S^2} \right]$$

Formel für:

Kuder-Richardson 20 (K-R20)

$$r = \frac{k}{k-1} \left[ 1 - \frac{\sum pq}{S^2} \right]$$

## ANHANG B

### Beispiel für einen Subtest mit Bewertungsvorgabe

Es handelt sich um die Prüfung *First Certificate in English* (FCE) und hier um den Subtest „*Use of English*“ (*FCE Paper 3*) von Juni 2004.

Der Subtest auf den nächsten Seiten besteht aus fünf itembasierten Aufgaben. Diesen Aufgaben folgen der Lösungsschlüssel und die Bewertungsvorgabe für diesen Subtest.

Teil 1 wird vom Computer ausgewertet.

Die Teile 2, 3, 4 und 5 werden durch Hilfskräfte ausgewertet.

Zu folgenden Punkten werden den Hilfskräften Anleitungen gegeben:

- welche Lösungen akzeptabel sind
- ob zusammengezogene Formen wie „haven't“ oder „don't“ akzeptabel sind
- ob eine korrekte Rechtschreibung verlangt wird
- der Punktwert für jedes Item

Die Hilfskräfte müssen der Bewertungsvorgabe strikt folgen und in Zweifelsfällen den Gruppenleiter fragen.

**SAMPLE PAPER: PAPER 3 USE OF ENGLISH****PAPER 3 Use of English****Part 1**

For questions 1-15, read the text below and decide which answer (A, B, C or D) best fits each space.

There is an example at the beginning (0).

Mark your answers on the separate answer sheet.

**Example:**

0      A consider      B know      C call      D label

|   |   |   |   |   |
|---|---|---|---|---|
| 0 | A | B | C | D |
|   |   |   |   |   |

**SHOPPING MALLS**

Victor Gruen, an American architect, revolutionised shopping in the 1950s by creating the type of shopping centre that we now (0) ..... a shopping mall.

Gruen's (1) ..... was to provide a pleasant shopping environment in the suburbs. This (2) ..... shutting out the noise of the city environment and also enabling people to shop in all kinds of weather. He (3) ..... on using building designs that he knew people would feel (4) ..... with, but placed them in landscaped 'streets' that were entirely enclosed and often covered with a curved glass roof. This was done to (5) ..... some of the older shopping arcades of city centres, but (6) ..... these housed only small speciality shops, Gruen's shopping malls were on a much grander (7) ..... .

Access to the whole shopping mall was gained by using the main doors, which (8) ..... the Shopping 'streets' from the parking (9) ..... outside. As there was no need to (10) ..... out bad weather, shops no longer needed windows and doors, and people could wander (11) ..... from shop to shop. The space(12) ..... to build a shopping mall and its vast car parks can usually only be found in the suburbs or on the (13) ..... of the city. In many cities, shopping malls now (14) ..... much more than just shops; cinemas, restaurants and other forms of entertainment are also (15) ..... in popularity.

- 
- |    |                       |                    |                      |                      |
|----|-----------------------|--------------------|----------------------|----------------------|
| 1  | <b>A</b> direction    | <b>B</b> aim       | <b>C</b> search      | <b>D</b> view        |
| 2  | <b>A</b> resulted     | <b>B</b> sought    | <b>C</b> intended    | <b>D</b> meant       |
| 3  | <b>A</b> insisted     | <b>B</b> demanded  | <b>C</b> requested   | <b>D</b> emphasised  |
| 4  | <b>A</b> favourable   | <b>B</b> agreeable | <b>C</b> comfortable | <b>D</b> enviable    |
| 5  | <b>A</b> model        | <b>B</b> imitate   | <b>C</b> repeat      | <b>D</b> shadow      |
| 6  | <b>A</b> while        | <b>B</b> even as   | <b>C</b> besides     | <b>D</b> in spite of |
| 7  | <b>A</b> measure      | <b>B</b> height    | <b>C</b> size        | <b>D</b> scale       |
| 8  | <b>A</b> disconnected | <b>B</b> withdrew  | <b>C</b> separated   | <b>D</b> parted      |
| 9  | <b>A</b> strips       | <b>B</b> lines     | <b>C</b> areas       | <b>D</b> plots       |
| 10 | <b>A</b> hold         | <b>B</b> get       | <b>C</b> stay        | <b>D</b> keep        |
| 11 | <b>A</b> freely       | <b>B</b> loosely   | <b>C</b> simply      | <b>D</b> entirely    |
| 12 | <b>A</b> obliged      | <b>B</b> required  | <b>C</b> desired     | <b>D</b> expected    |
| 13 | <b>A</b> side         | <b>B</b> limit     | <b>C</b> edge        | <b>D</b> extent      |
| 14 | <b>A</b> contain      | <b>B</b> concern   | <b>C</b> consist     | <b>D</b> compose     |
| 15 | <b>A</b> becoming     | <b>B</b> growing   | <b>C</b> raising     | <b>D</b> advancing   |

**Part 2**

For questions 16-30, read the text below and think of the word which best fits each space. Use only **one** word in each space. There is an example at the beginning (0).

Write your answers on the separate answer sheet.

Example:

|   |       |
|---|-------|
| 0 | among |
|---|-------|

---

## DICTIONARIES

Dictionaries are (0) ..... the most important tools of self-education. (16) ..... Samuel Johnson wrote his influential English dictionary in the eighteenth century, the work kept him busy for seven years. At the end of that period, he (17) ..... written the meanings of over forty thousand words. Most modern dictionaries require a (18) ..... deal less time and effort to write because writers often use earlier dictionaries (19) ..... a source of reference.

(20) ..... it is possible for one person to write a dictionary alone, most dictionaries are team efforts. First of all, the writers, or lexicographers, draw up the rules that will guide their writing. For example, if a word has two meanings, they (21) ..... to agree about which order to put them (22) ..... . However, for much of the time, team members are able to work independently of (23) ..... other, on different parts of the dictionary.

(24) ..... one time, the starting point for deciding on those words to include used to be the lexicographer's own knowledge. These days, some teams (25) ..... use of a large collection of examples of (26) ..... only writing but also everyday speech, which is known as a *corpus*. Teams also refer (27) ..... books and articles about language as (28) ..... as asking experts in particular subjects about the more specialised words. Finally, ordinary people are asked to say what they think about the (29) ..... the words are defined and (30) ..... they find the examples provided helpful or not.

**Part 3**

For questions **31-40**, complete the second sentence so that it has a similar meaning to the first sentence, using the word given. **Do not change the word given**. You must use between **two** and **five** words, including the word given. Here is an example (**0**).

**Example:**

**0** You must do exactly what the manager tells you.

**Carry**

You must ..... instructions exactly.

The space can be filled by the words 'carry out the manager's', so you write:

|          |                         |
|----------|-------------------------|
| <b>0</b> | carry out the manager's |
|----------|-------------------------|

Write **only** the missing words **on the separate answer sheet**.

**31** Today's meeting is postponed and it will be held next week.

**put**

Today's meeting has ..... until next week.

**32** Unfortunately, Kim couldn't go to the cinema because she didn't have any money.

**able**

If Kim had had some money, she ..... go to the cinema.

**33** According to the report, the driver of the car was a policeman.

**being**

According to the report, the ..... by a policeman.

**34** Nobody spoke for about five minutes.

**before**

It was about five minutes ..... anything.

**35** Mr Johnson continued to get up at 6.30 even after he retired.

**carried**

Mr Johnson ..... at 6.30 even after he retired.

**36** I prefer eating sandwiches to a cooked lunch.

**rather**

I ..... sandwiches than a cooked lunch.

37 'I'm sorry I behaved so badly,' said George.

**apologised**

George ..... so badly.

38 There's no chance of Jenny getting here on time.

**possible**

It won't be ..... here on time.

39 'We really don't need to leave early,' said Elena.

**point**

'There's really ..... early,' said Elena.

40 Cars couldn't get onto the motorway because of an accident.

**prevented**

An accident ..... onto the motorway.

**Part 4**

For questions 41-55, read the text below and look carefully at each line. Some of the lines are correct, and some have a word which should not be there.

If a line is correct, put a tick (✓) by the number on the separate answer sheet. If a line has a word which should not be there, write the word on the separate answer sheet. There are two examples at the beginning (0 and 00).

|   |   |
|---|---|
| 0 | ✓ |
|---|---|

|    |      |
|----|------|
| 00 | more |
|----|------|

**Examples:**

---

**MY GRANDMOTHER**

- 0 The person I am going to write about is my grandmother. She is a  
00 lively lady of more seventy-five years, although to look at her you would  
41 think she is ten years younger aged. Now she lives in the city  
42 and my brothers and I are often go to visit her at weekends. She  
43 always gives to us coffee and some of her delicious home-made cake  
44 and then we sit and listen to her so fascinating stories of the days  
45 when she was being a girl. At that time, she lived deep in the  
46 countryside in a place where there were very few cars or buses  
47 and she was used to have to walk ten kilometres just to get to  
48 school. There were no televisions, so she would amused herself  
49 by playing outside with her friends until that it got dark. In  
50 winter, when it was too completely dark, wet or cold to go outside,  
51 she read books that she had borrowed from the local library and  
52 learned a lot about places beyond the village she lived in there.  
53 Eventually, by studying hard, she became a teacher and started the  
54 work in a school in the city, and it was there that she met my  
55 grandfather, who worked as an engineer in a large factory nearby it.

**Part 5**

For questions **56-65**, read the text below. Use the word given in capitals at the end of each line to form a word that fits in the space in the **same** line. There is an example at the beginning **(0)**.

Write your answers on the separate answer sheet.

Example:

|   |          |
|---|----------|
| 0 | tropical |
|---|----------|

---

### ISLAND IN THE SUN

With its **(0)** ..... sunshine and warm welcome, this island is **TROPIC**  
 hard to beat as a holiday destination. The **(56)** ..... west coast is the **DELIGHT**  
 perfect **(57)** ..... for lovers of water sports and sunbathing. There is **CHOOSE**  
 also an **(58)** ..... selection of restaurants, where the local seafood **IMPRESS**  
 is **(59)** ..... recommended, and beaches of fine white sand face on to the **HIGH**  
 calm Caribbean Sea.

Other **(60)** ..... on the island include underwater trips in a submarine and **ATTRACT**  
 a jazz festival held **(61)** ..... , early in January. It is also worth travelling **ANNUAL**  
 along the wild east coast, which is often **(62)** ..... as it faces the Atlantic **STORM**  
 Ocean. This makes the coast **(63)** ..... for swimming, unlike the calmer **SUITABLE**  
 beaches on the west coast. Car hire is **(64)** ..... arranged here, and **EASY**  
 there is a good road system, with a very **(65)** ..... bus service to take **RELY**  
 you around the island.

**ANSWER KEY: Paper 3: Use of English****Part 1**

1 B

2 D

3 A

4 C

5 B

6 A

7 D

8 C

9 C

10 D

11 A

12 B

13 C

14 A

15 B

**Part 2**

16 When

17 had

18 great

19 as

20 (A)though

21 have

22 in

23 each

24 At

25 make

26 not

27 to

28 well

29 way

30 whether

**Part 3**31 been | **put off**32 would have been | **able to**33 car was **being** | driven34 **before** anybody | said35 **carried on** | getting up36 would **rather** | have37 **apologised** for | behaving38 **possible** for | Jenny to get39 no **point** in | us/our leaving40 **prevented** cars | from getting

## ANHANG C

### Beispiel für einen Subtest zum Schreiben mit Bewertungsanleitungen und Musterkorrekturen

#### **First Certificate in English, Subtest 2**

Der folgende Subtest zum Schreiben wird von geschulten Beurteilern mit verschiedenen Bewertungsanleitungen bewertet, die im Anschluss an die Aufgaben vorgestellt werden.

Die schriftlichen Arbeiten der Kandidaten, die im Anschluss an die Bewertungsanleitungen gezeigt werden, wurden vom *Chief Examiner* für die Musterkorrektur ausgewählt und sollen typische Leistungen auf bestimmten Stufen exemplifizieren. Die Trainings- und Standardisierungsverfahren für die Beurteiler werden im Kapitel „4 a) Tests zum Schreiben“ in diesem Modul beschrieben.

**SAMPLE PAPER: PAPER TWO COMPOSITION****Instructions to candidates**

Answer Question 1 and **one** of the Questions 2-5.  
Each question in this paper carries equal marks.

**Part 1**

You **must** answer this question.

**1** You are a student at a college in England and you have seen the article below in the college newsletter. Read the article and the notes you have made. Then write a letter to David Brown, the College Director, using **all** your notes.

**A generous gift for our college!**

John Maitland, a former student of our college and now a millionaire businessman, has given the college a large sum of money to improve the college facilities. He wants the money to be used to provide something the students need.

*Great!*

David Brown, our College Director, is concerned that students do not do enough sport and has therefore suggested that the money should be used to build a swimming pool. The pool would replace the college garden.

*Not true...*

*Not necessary because...*

*No! Explain why students like the garden*

*Suggest...*

Building work would start next year, so there is still time for students to suggest other ways of spending the money.

Write a **letter** of between **120** and **180** words in an appropriate style on the opposite page. Do not write any postal addresses.

**Part 2**

Write an answer to **one** of the questions 2-5 in this part. Write your answer in **120-180** words in an appropriate style on the opposite page. Put the question number in the box at the top of page 5.

---

**2** A British TV company is thinking of making a film about life in your area and has asked you to give them some information. Write a report describing the advantages of living in your area and saying how the area might change in the future.

Write your **report**.

**3** You see this announcement in an English language magazine.

**COLOUR**

- What is your favourite colour and why?
- Why is colour so important in our lives?

Write us an article answering these questions.  
The best article will be published in the magazine.

Write your **article**.

**4** You have had a class discussion about sport. Now your English teacher has asked you to write a composition discussing the following question:

*Should professional footballers be paid more than doctors?*

Write your **composition**.

**5** Answer **one** of the following two questions based on your reading of **one** of these set books. Write the letter **(a)** or **(b)** as well as the number **5** in the question box, and the **title** of the book next to the box. Your answer **MUST** be about one of the books below.

*Round the World in Eighty Days* – Jules Verne

*Pride and Prejudice* – Jane Austen

*The Prisoner of Zenda* – Anthony Hope

*Deadlock* – Sara Paretsky

*Ghost Stories* – retold by Rosemary Border

**Either (a)** An international magazine is looking for articles about brave characters in books. Write an **article** describing a character and explaining why you think that character shows courage in the book or short story you have read.

**Or (b)** 'At the end of a story, all the reader's questions are answered.' How true is this of the book or short story you have read? Write a **composition** explaining your views with reference to the book or short story you have read.

## FIRST CERTIFICATE IN ENGLISH: SUBTEST 2, BEWERTUNG

Der Subtest zum Schreiben besteht aus zwei gleich gewichteten Teilen. Die Aufgabe in Teil 1 ist für alle Kandidaten verpflichtend und besteht darin, einen formellen Brief als Antwort auf einen gegebenen Input zu schreiben. In Teil 2 wählen die Kandidaten aus vier Aufgaben eine Aufgabe aus. Die Kandidaten müssen also insgesamt nur zwei Aufgaben bearbeiten. Wenn ein Kandidat die Aufgabe in Teil 1 nicht bearbeitet, erhält er für diesen Teil einen Punktwert von 0. Falls in Teil 2 zwei Aufgaben bearbeitet werden, werden beide bewertet und der höhere Punktwert wird vergeben.

### FOKUS DER BEWERTUNG

Wie im Folgenden erläutert, ist der Fokus der Aufgabe in beiden Teilen des Subtests verschieden. Die Prüfer bewerten beide Teile mithilfe der aufgabenunabhängigen Bewertungsanleitung, aber jede Aufgabe zusätzlich noch mit der aufgabenspezifischen Bewertungsanleitung, wenn die Leistung befriedigend oder besser ist (d. h. auf der Leistungsstufe 3 oder höher).

Teil 1. Bei der Bewertung des Briefes liegt der Fokus auf Vollständigkeit des Inhalts, Korrektheit der Sprache und Angemessenheit des Registers und des Formats im Hinblick auf den Leser. Das Spektrum der sprachlichen Mittel wird durch die Merkmale der Aufgabe bestimmt. Die Kandidaten können wichtige Wörter aus dem Input in ihren Brief übernehmen. Längere Formulierungen zu übernehmen, führt jedoch zu Punktabzug.

Teil 2. Die Aufgabentypen und Themen in Teil 2 erlauben den Kandidaten eine größere Eigeninitiative. Bei der Bewertung der vom Kandidaten gewählten Aufgabe liegt die Betonung auf der Breite des Wortschatzes und der Strukturen sowie auf der Korrektheit. Es gibt einen größeren Spielraum, was den Inhalt und den Adressaten betrifft. Die Textsorten Artikel und Bericht haben kein streng vorgeschriebenes Format, dennoch werden für eine angemessene Präsentation Punkte vergeben.

### BEWERTUNGSVERFAHREN

Jede schriftliche Arbeit wird auf einer Stufe zwischen 0 und 5 bewertet; die jeder Stufe entsprechenden Leistungen werden in der aufgabenunabhängigen Bewertungsanleitung beschrieben. Jede der fünf Leistungsstufen wird in drei Unterstufen unterteilt, zum Beispiel 3.1, 3.2, 3.3 (um die Leistung auf einer bestimmten Stufe zusätzlich noch zu qualifizieren). Bei der Bewertung mit der aufgabenspezifischen Bewertungsanleitung erfolgen Punktabzüge, wenn der Kandidat die dort genannten Anforderungen nicht erfüllt.

### LÄNGE

Es werden 120-180 Wörter verlangt.

Wird dieser Umfang unterschritten (50-100 Wörter), wird die Leistung gemäß den Leistungsstufen 1 und 2 der aufgabenunabhängigen Bewertungsanleitung bewertet. Arbeiten mit weniger als 50 Wörtern werden mit 0 bewertet.

Wird der Umfang überschritten (mehr als 200 Wörter), liest der Beurteiler zunächst die ganze Arbeit. Beim zweiten Lesen zieht er eine Linie dort, wo der korrekte Umfang erreicht wurde und bewertet dann nur den Text oberhalb dieser Linie.

### RECHTSCHREIBUNG UND ZEICHENSETZUNG

Beide betreffen das Kriterium Korrektheit. Sie werden je nach dem Grad bewertet, in dem sie die Kommunikation beeinträchtigen oder verhindern. Amerikanische Schreibweise wird geduldet, vorausgesetzt sie wird konsistent angewendet.

### ABSÄTZE

Dies betrifft das Kriterium Aufbau, siehe auch „Aufbau und Kohäsion“ in der aufgabenspezifischen Bewertungsanleitung.

### HANDSCHRIFT

Wenn die Handschrift des Kandidaten die Kommunikation beeinträchtigt, ohne sie zu verhindern, wird die Leistung eine Stufe niedriger bewertet. Unlesbarkeit wird mit einem Punktwert von 0 bewertet.

### Aufgabenunabhängige Bewertungsanleitung

|   |  |
|---|--|
| 5 | <p>Aufgabe voll erfüllt.</p> <ul style="list-style-type: none"> <li>▪ Alle Inhaltspunkte werden in angemessenem Umfang behandelt.</li> <li>▪ Großes Spektrum der Strukturen und des Wortschatzes gemäß der Aufgabe.</li> <li>▪ Minimale Fehler, vielleicht weil der Kandidat/die Kandidatin risikofreudig oder ehrgeizig ist; gut entwickelte Beherrschung der Sprache.</li> <li>▪ Effektiver Aufbau der Gedanken mit einer Vielfalt verbindender Elemente.</li> <li>▪ Register und Format durchgehend der Schreibabsicht und dem Leser angemessen.</li> </ul> <p>Erzielt voll die gewünschte Wirkung auf den Leser.</p>   |
| 4 | <p>Gute Erfüllung der Aufgabe.</p> <ul style="list-style-type: none"> <li>▪ Alle wichtigen Inhaltspunkte werden behandelt; ein oder zwei kleinere Punkte werden möglicherweise nicht aufgegriffen.</li> <li>▪ Völlig angemessenes Spektrum der Strukturen und des Wortschatzes gemäß der Aufgabe.</li> <li>▪ Im Allgemeinen korrekt; Fehler hauptsächlich dort, wo versucht wird, komplexere Sprache zu verwenden.</li> <li>▪ Klarer Aufbau der Gedanken mit angemessenen verbindenden Elementen.</li> <li>▪ Register und Format im Allgemeinen der Schreibabsicht und dem Leser angemessen.</li> </ul> <p>Erzielt die gewünschte Wirkung auf den Leser.</p>   |
| 3 | <p>Ausreichende Erfüllung der Aufgabe.</p> <ul style="list-style-type: none"> <li>▪ Alle wichtigen Inhaltspunkte werden behandelt; einige kleinere Auslassungen.</li> <li>▪ Angemessenes Spektrum der Strukturen und des Wortschatzes, die die Anforderungen der Aufgabe erfüllen.</li> <li>▪ Eine Reihe von Fehlern ist möglich, sie behindern jedoch nicht die Kommunikation.</li> <li>▪ Angemessener Aufbau der Gedanken mit einfachen Verbindungen.</li> <li>▪ Ausreichender, wenn auch nicht immer erfolgreicher Versuch, ein der Schreibabsicht und dem Leser angemessenes Register und Format zu verwenden.</li> </ul> <p>Erzielt im Großen und Ganzen die gewünschte Wirkung auf den Leser.</p>          |
| 2 | <p>Versuch, die Aufgabe zu erfüllen, der aber nicht angemessen gelingt.</p> <ul style="list-style-type: none"> <li>▪ Einige wichtige Inhaltspunkte werden unzureichend oder gar nicht behandelt und/oder es gibt irrelevante Äußerungen.</li> <li>▪ Begrenztes Spektrum der Strukturen und des Wortschatzes.</li> <li>▪ Eine Reihe von Fehlern, die den Leser ablenken und zeitweise die Kommunikation behindern können.</li> <li>▪ Nicht angemessener Aufbau der Gedanken; nur selten Verwendung von verbindenden Elementen.</li> <li>▪ Kein erfolgreicher/konsistenter Versuch, ein angemessenes Register und Format zu verwenden.</li> </ul> <p>Die Aussage wird dem Leser nicht klar genug kommuniziert.</p> |
| 1 | <p>Schwacher Versuch, die Aufgabe zu erfüllen.</p> <ul style="list-style-type: none"> <li>▪ Erhebliche inhaltliche Auslassungen und/oder häufig irrelevante Inhalte, möglicherweise wegen Nicht-Verstehen der Aufgabe bzw. der Anweisung.</li> <li>▪ Geringes Spektrum an Wortschatz und Strukturen.</li> <li>▪ Häufige Fehler, die die Kommunikation behindern; kaum entwickelte Beherrschung der Sprache.</li> <li>▪ Kein Aufbau der Gedanken, keine verbindenden Elemente.</li> <li>▪ Kein oder nur geringes Bewusstsein für Register und Format.</li> </ul> <p>Die Äußerung hat eine sehr negative Wirkung auf den Leser.</p>  |
| 0 | <p>Der Kandidat erreicht nichts. Zu wenig Sprache (unter 50 Wörtern), um sie zu bewerten, irrelevanter Inhalt oder unleserliche Handschrift.</p>   |

Diese Bewertungsanleitung sollte vor dem Hintergrund einer typischen Leistung auf dem Niveau des *First Certificate in English* (B2) interpretiert und zusammen mit der aufgabenspezifischen Bewertungsanleitung angewendet werden. Innerhalb jeder Leistungsstufe (engl.: *band*) können drei Abstufungen vorgenommen werden. Eine Leistung mit einem Testwert von 3.1 wäre eine auf dem Niveau des FCE befriedigende Leistung und ein Testwert von 5.3 wäre eine hervorragende Leistung auf dem Niveau des FCE, wobei die Arbeit nicht völlig fehlerfrei sein muss.

## **AUFGABENSPEZIFISCHE BEWERTUNGSANLEITUNG FÜR DAS FCE FÜR DIE LEISTUNGSSTUFEN 3.1 UND HÖHER**

### **AUFGABE 1**

#### **INHALT**

Der Brief muss alle folgenden Punkte enthalten:

- 1) ausdrucksvolle positive Reaktion auf das Geschenk.
- 2) Hinweis darauf, dass die Studenten genug Sport treiben und/oder ausreichende Sportanlagen zur Verfügung stehen.
- 3) Hinweis darauf, dass ein Schwimmbad nicht gewünscht wird.
- 4) Hinweis darauf, dass das Schwimmbad nicht den Garten ersetzen soll und/oder Begründung, warum die Studenten den Garten mögen.
- 5) Vorschlag, wie man das Geld anderweitig verwenden könnte.

#### **AUFBAU UND KOHÄSION**

Klarer Aufbau der Argumentation mit angemessenen Absätzen und entsprechenden verbindenden Elementen. Einleitung und Schluss sind angemessen.

#### **ANGEMESSENHEIT DES REGISTERS UND DES FORMATS**

Kann formal oder informell sein, vorausgesetzt es ist konsistent.

#### **SPEKTRUM DER SPRACHLICHEN MITTEL**

Sprache, um eine ausdrucksvolle positive Reaktion auszudrücken, um Vorschläge zu machen, zu erklären, zu beschreiben und um zu widersprechen oder zu kommentieren.

#### **ADRESSAT**

Er wäre informiert.

|   |   |
|---|---|
| <p><b>AUFGABE 2</b></p> <p><b>INHALT</b><br/>Der Bericht sollte die Vorteile der Wohngegend beschreiben, in der der Verfasser wohnt, sowie mögliche zukünftige Veränderungen. Es ist nicht erforderlich, die Gegend zu benennen.</p> <p><b>SPEKTRUM DER SPRACHLICHEN MITTEL</b><br/>Sprache, um zu beschreiben, Meinungen auszudrücken und zu erklären.</p> <p><b>AUFBAU UND KOHÄSION</b><br/>Der Bericht sollte klar aufgebaut sein mit einer Einleitung und einem Schluss. Zwischenüberschriften sind von Vorteil.</p> <p><b>ANGEMESSENHEIT DES REGISTERS UND DES FORMATS</b><br/>Register kann neutral oder formell sein, muss aber konsistent bleiben. Das formale Layout eines Berichts wird nicht gefordert.</p> <p><b>ADRESSAT</b><br/>Er wäre informiert.</p> |   |
| <p><b>AUFGABE 3</b></p> <p><b>INHALT</b><br/>Der Artikel sollte die Lieblingsfarbe(n) beschreiben und darstellen, warum Farbe wichtig ist.</p> <p><b>SPEKTRUM DER SPRACHLICHEN MITTEL</b><br/>Sprache, um zu erklären, zu beschreiben und Meinungen auszudrücken.</p> <p><b>AUFBAU UND KOHÄSION</b><br/>Klarer Aufbau der Gedanken mit angemessenen Absätzen und verbindenden Elementen.</p> <p><b>ANGEMESSENHEIT DES REGISTERS UND DES FORMATS</b><br/>Jedes Register, vorausgesetzt es wird durchgehend verwendet.</p> <p><b>ADRESSAT</b><br/>Er wäre informiert.</p>   | <p><b>AUFGABE 4</b></p> <p><b>INHALT</b><br/>Der Aufsatz sollte die gestellte Frage diskutieren. Es ist nicht erforderlich, die Frage mit einem definitiven Ja oder Nein zu beantworten.</p> <p><b>SPEKTRUM DER SPRACHLICHEN MITTEL</b><br/>Sprache, um zu diskutieren und Meinungen auszudrücken.</p> <p><b>AUFBAU UND KOHÄSION</b><br/>Klarer Aufbau der Gedanken mit angemessenen Absätzen und verbindenden Elementen.</p> <p><b>ANGEMESSENHEIT DES REGISTERS UND DES FORMATS</b><br/>Neutrales Register eines Aufsatzes.</p> <p><b>ADRESSAT</b><br/>Er wäre informiert.</p> |

|  |   |
|--|---|
| <p><b>AUFGABE 5a</b></p> <p><b>INHALT</b><br/>Der Schreiber sollte einen oder zwei Charakter(e) aus dem Buch beschreiben und erklären, auf welche Weise sie Mut zeigen.</p> <p><b>SPEKTRUM DER SPRACHLICHEN MITTEL</b><br/>Sprache, um Meinungen auszudrücken, zu erklären und zu beschreiben.</p> <p><b>AUFBAU UND KOHÄSION</b><br/>Klarer Aufbau der Gedanken mit angemessenen Absätzen und verbindenden Elementen.</p> <p><b>ANGEMESSENHEIT DES REGISTERS UND DES FORMATS</b><br/>Jedes Register, vorausgesetzt es wird durchgehend verwendet.</p> <p><b>ADRESSAT</b><br/>Er wäre informiert.</p> | <p><b>AUFGABE 5b</b></p> <p><b>INHALT</b><br/>Der Aufsatz sollte mit der gemachten Aussage übereinstimmen oder ihr widersprechen.</p> <p><b>SPEKTRUM DER SPRACHLICHEN MITTEL</b><br/>Sprache, um zu erklären und Meinungen auszudrücken.</p> <p><b>AUFBAU UND KOHÄSION</b><br/>Klarer Aufbau der Gedanken mit angemessenen Absätzen und verbindenden Elementen.</p> <p><b>ANGEMESSENHEIT DES REGISTERS UND DES FORMATS</b><br/>Neutrales Register eines Aufsatzes.</p> <p><b>ADRESSAT</b><br/>Er wäre informiert.</p> |
|--|---|

**BEISPIELE FÜR KANDIDATENLEISTUNGEN MIT BEWERTUNG**

Die Anmerkungen zur Bewertung beziehen sich auf die aufgabenunabhängige Bewertungsanleitung auf Seite 34

**Question 1****Script A**

*Dear Mr Brown,*

*I am writing in response to the note in which you inform us about this fantastic news that the college will receive new funds.*

*I would like to react to your intention to spend this money on building a swimming pool. Indeed, after many conversations with other students, we unfortunately disagree with this idea because we feel that we practise enough sports with the equipments already provided such as the gymnasium, the football pitch and the tennis courts. Moreover, just five minutes walk from the college, there is an olympic swimming pool which offers very interesting fares for the students.*

*Therefore, I would like to point at the fact that students absolutely want to preserve 'their' garden because it allows us to have fresh air during our breaks, this is a place where we can relax. So we would like to preserve a little space of nature, which could be important for our concentration.*

*That is why I would like to suggest you to use the money on buying new computers, which we desperately need, or buying audio equipment for the Foreign Languages Department for instance.*

*I look forward to hear from.*

*Yours sincerely,*

**Question 1****Script A**

**INHALT:** alle Punkte in angemessenem Umfang berücksichtigt

**KORREKTHEIT:** kleine Fehler, die auf Risikofreudigkeit zurückzuführen sein können

**SPEKTRUM SPRACHLICHER MITTEL:** breites Spektrum

**AUFBAU UND KOHÄSION:** guter Aufbau

**ANGEMESSENHEIT DES REGISTERS UND DES FORMATS:** durchgehend angemessen

**ADRESSAT:** wäre vollkommen informiert

**Stufe 5**

**Question 2****Script B**

*About life in my area is Malaysia. I born in malaysia. In the malaysia the wherther is very hot and sunshine.*

*In my country got many nice place and the nice food in there. Seaside is very popular in my country. There very shinary people like swimming and take boat go around the sea. In the night there walk in the beach. In my country have the night market. In the market many thing to sell. There have food. Clote, drink, CD movie, muzie and many thing there would many people eat in there.*

*In the future my country would change.*

**Question 2****Script B**

**INHALT:** schwacher Versuch, die Aufgabe zu erfüllen

**KORREKTHEIT:** häufige Fehler behindern die Kommunikation, keine ausreichende Sprachbeherrschung

**SPEKTRUM SPRACHLICHER MITTEL:** eingeschränktes Spektrum

**AUFBAU UND KOHÄSION:** Versuch einer Struktur erkennbar

**ANGEMESSENHEIT DES REGISTERS UND DES FORMATS:** angemessen

**ADRESSAT:** sehr negative Wirkung auf den Leser

**Stufe 1**

**Question 3****Script C****COLOUR IS LIFE**

*To my opinion, nothing is more beautiful than a rainbow: it's 'the result of the fight between sun and rain'. In other words, colour is the way to express our feelings.*

*My favourite colour depends on my life, it changes as changes my life. But if I had to choose I would say blue. Blue because it reminds me the sea. I feel lively when I am by the sea, so I guess that being surrounded by blue helps me feeling lively.*

*But I have noticed that people, included me, have a tendancy in wearing bright colours when they are happy, and dark colours when they are not. Why?*

*I think colours can be qualified as a language. You can guess people personality by looking at their clothes, car, or else. But colours can influence our mood too. If we were living in a black and white world, people wouldn't be happy I think, everybody would be depressed. Colour gives life to anything. And there is no doubt that people feel better surrounded by colour.*

*For a happy world, we need colour in our lives, because colour is life.*

**Question 3****Script C**

**INHALT:** gute Entwicklung der Aufgabe, führt zu einem geeigneten Schluss

**KORREKTHEIT:** Fehler lenken den Leser nicht ab, obwohl einige Flüchtigkeitsfehler enthalten sind

**SPEKTRUM SPRACHLICHER MITTEL:** gute idiomatische Strukturen, Gebrauch des Konditionals und Spektrum des Wortschatzes

**AUFBAU UND KOHÄSION:** gut aufgebaut

**ANGEMESSENHEIT DES REGISTERS UND DES FORMATS:** angemessen

**ADRESSAT:** wäre gut informiert

**Stufe 4**

**Question 4****Script D***Entertainment and Reality*

*Although I've been an apassionate of football since i was a child, I think that footballers should not be paid more than doctors.*

*Firstly, because I don't consider football as a profession. Football is only an enterteinment where footballers perfom. In few words, it's like actors and actresses in the movies. The doctors instead are saving lifes. What's more important or remarkable than that? They are not valuated and also they are accussed of negligency. Sometimes. People should think about what is worth to be paid as deserves and not. I think doctors are real.*

**Question 4****Script D**

**INHALT:** Aufgabe angemessen bewältigt, obwohl die erforderliche Wortzahl leicht unterschritten wurde

**KORREKTHEIT:** einige irritierende Fehler und einige, die das Verständnis beeinträchtigen

**SPEKTRUM SPRACHLICHER MITTEL:** begrenzt

**AUFBAU UND KOHÄSION:** kaum vorhanden

**ANGEMESSENHEIT DES REGISTERS UND DES FORMATS:** keine einheitliche Ausführung

**ADRESSAT:** Information nicht klar vermittelt

**Stufe 2**

**Question 5a****Script E***'Round The World In Eighty Days'*

*One of the brave characters in book 'Round The World In Eighty Days' written by Jules Verne is Mr. Phileas Fogg. He makes a bet with his friends from The Reform Club that he can travel round the world in eighty days. He can loose everything if he doesn't do so. Don't you think it is a good start how to prove he's such a brave English gentleman? I bet you do.*

*It is a colourful race through Europe, Asia and America follows. They experience a lot of unexpected situations and delays. Mr Phileas Fogg and his French servant Passepartout face a danger travelling through India.*

*Let me tell you about one of the story I have read. As they travel through India meeting Indians priests who want to burn a beautyfull princess Aouda. They make a plan how to save her life. If Mr Fogg is not a brave man, he'll continue his journey without stoping and he wouldn't even think about saving Aouda. Of coure they were successful.*

*I can strongly recommended this book to anyone who is looking for a remarkable read.*

**Question 5a****Script E**

**INHALT:** gute Erfüllung der Aufgabe

**KORREKTHEIT:** einige Fehler, die aber nicht das Verständnis beeinträchtigen, besonders in der 2. Hälfte

**SPEKTRUM SPRACHLICHER MITTEL:** angemessen

**AUFBAU UND KOHÄSION:** angemessener Aufbau

**ANGEMESSENHEIT DES REGISTERS UND DES FORMATS:** ausreichend

**ADRESSAT:** erzielt die gewünschte Wirkung beim Leser

**Stufe 3**