**Appendix B**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Assessment situations | Agree | Disagree |
| 1 | Steven finds it difficult to actually read the text due to sloppy handwriting. In his opinion, this impairs the reading experience, and he chooses to take it into consideration when assessing the scale for *Writer-reader-interaction*. | 1 | 0 |
| 2 | David notices that a student text suffers from poor text structure. On the specific scale for assessing structure, David is convinced that this text merits the lowest mark; however, it is obvious that the poor text structure is a result of ‘learning-in-progress’. David decides to acknowledge this and gives the text a higher mark. | 1 | 0 |
| 3 | Sheryl consults several colleagues in the NPR when she finds that the guideline is unreasonable in one specific area. Supported by her colleagues, she decides that she is right, and she therefore chooses to ignore the guideline in that specific area. | 1 | 0 |
| 4 | John has just read a text that ‘hit him’. Even though he can’t defend giving the text a top score on every scale using the assessment resources, he still does because the text touched him so tremendously. | 1 | 0 |
| 5 | Martha and Peter allocate much of their time to pair assessments to discuss issues related to writing and writing assessments that they find to be of special interest. They appear to agree on many points, and at the same time, they receive help from each other to improve their classroom practices. Since they appear to agree on writing and writing assessments, they do not spend much time discussing the actual ratings they have carried out. | 1 | 0 |
| 6 | Richard gives low text scores specifically on *Communication*, *Content*, *Text structure* and *Language use*. To avoid discouraging the student, Richard considers giving higher text scores on *Morphology, spelling and punctuation* than he believes the text merits. He hesitates but decides not to give the higher scores. | 0 | 1 |
| 7 | Stephen struggles with the assessment of a specific text. He believes that it is the best text in the stack; however, the text can hardly be said to represent the genre the assignment requires. Ultimately, Stephen decides to assess the text based on its inherent qualities. | 1 | 0 |
| 8 | Joe finds one of the texts in the stack to be problematic. The content is relevant, but Joe believes that the student has not taken the assignment seriously. The arguments in the text appear to be unsound. Consequently, Joe decides to give the text a low mark. | 1 | 0 |
| 9 | Lisbeth scores a text that clearly communicates the topic to the reader; however, due to what she believes is a poor writing prompt, the Writer-reader-interaction is not entirely relevant. She scores the text ‘M2’ (mastery level 2) even though she believes the text merits ‘M4’ on Writer-reader-interaction. | 0 | 1 |
| 10 | Joy finds it difficult to conceive of what to expect from the students’ texts. She is not capable of this until she has assessed half the stack; however, at this point, she finds it unreasonable to go back and check earlier scores due to the time limit. | 1 | 0 |
| 11 | Roger does not assess all texts in a specific order. He begins with *Morphology, spelling and punctuation* when he reviews a poorly written text, and he begins with any of the other scales when he reviews a well-written text. | 1 | 0 |
| 12 | Melanie suspects that the text at hand is an ‘L2’ text. She finds serious mistakes related to morphology and spelling, but she believes the punctuation is slightly better than can be expected from an L2 student. Forced to choose between ‘M1’ and ‘M2’, Melanie places emphasis on the student’s competence related to punctuation and decides to give the student an ‘M2’ for this specific scale. | 1 | 0 |
| 13 | To score a text ‘M5’ on *Language use*, the text must include a complex and varied sentence structure. Yvette’s experience tells her that not even the best student texts reach such a level. To be able to use this level (‘M5’), Yvette chooses to ignore the criteria. | 1 | 0 |
| 14 | During one pair assessment session, Lucy and Annie constantly disagree about the scoring. Lucy nevertheless wishes to reach an agreement, and therefore she almost always allows herself to be convinced by Annie. The consequence is that the scores they assign almost always are scores that Annie suggests. | 1 | 0 |
| 15 | Linda believes that in general, the students’ mastery levels related to writing are low. She therefore believes that she, as a member of the National Assessment Panel, must be harsh to avert a further decrease in mastery. | 1 | 0 |
| 16 | Ellen normally scores as many texts as possible in one session. She finds this way of organising scoring effective, and she also believes she needs to read some texts ‘to get into it’. | 1 | 0 |
| 17 | Angelina reacts when she reads the guideline’s phrasing related to the *Writer-reader interaction*. Her professional experience tells her that the guideline is not accurate. She believes that it has been a mistake to include this formulation in the guideline. Angelina therefore chooses to ignore the phrasing in the guideline when assessing texts, and she also decides to inform the test developers about the weakness within the guideline. | 1 | 0 |
| 18 | Joanne reads a text that is very short and only includes a few sentences. The only punctuation used is a full stop, and this is done correctly. Since Joanne cannot find any mistakes related to punctuation in the text, she finds herself forced to give the text ‘M5’ on this specific scale. | 1 | 0 |
| 19 | Anna knows that interrater reliability is very demanding, if not impossible. This makes her believe that the most important thing the NPR’s work can lead to is to help NPR members develop critical text competence to enable them to justify their scores. | 1 | 0 |
| 20 | Willy checks his scores for the different scales for a given text. He now sees that the sum of his scores does not coincide with his holistic impression of the text. He actually believes the text is better than his analytic scoring shows; however, Willy decides not to adjust the scores to make them fit better with his holistic impression of the text. | 0 | 1 |
| 21 | Martha believes she must recognise that she cannot always give the correct mark and that she therefore must sometimes go back and reassess. | 0 | 1 |
| 22 | As a part of preparation for scoring, Kurt makes a ‘blind’ assessment of one of the benchmark texts. When he then compares his scores with the scores that follow the guideline, he discovers that he is more lenient than the guideline. He believes that the benchmark text is not in line with reality. Still, he decides to try to be harsher than he actually believes is accurate just to be in line with the assessment resources. | 0 | 1 |
| 23 | Bernie believes the text he is scoring is as expected (‘M3’) for most of the scales. He still feels that the text is better than most of the other texts he has already scored within the same band level. He therefore decides to give this text an ‘M4’ on most of the scales to mark the difference in quality. | 1 | 0 |