CHAPTER 4

FINDINGS AND DISCUSSION

4.1 Chapter Overview

This chapter presents the findings and discussion of the findings that would later be used to answer the research questions of the study. For the first phase of the study, that is the mapping out the ER mental model, the data collected through verbal protocol, interviews, journal and reflections was analyzed to gain insights on the participants' personal justification on the decision and action taken while doing the process of 'thinking aloud' in assessing writing. The findings are presented in tabulated forms as to show the frequency counts. Description and interpretation of the findings are provided after each table. From the findings, a mental model was constructed to represent how the expert raters cognitively assessed and gave quality feedback students' writing. The findings in phase one were later used as a scaffold in phase two, which is trialling out the ER mental model in the form of training sessions with a group of beginning teachers or novice teachers. The data collected through interviews from the participants in the second phase of the study were then used to answer the research question - How can the novice raters approximate the expert raters in assessing behaviour through training based on the expert rater mental model?

4.2 Research Findings

Verbal protocols obtained through the think-aloud method would be useful as an effective tool for understanding how cognitive skills and strategies change and develop (Robinson, 2001). This is so as verbal protocols can provide more complete information on raters' problem-solving strategies, and also a mean of assessing the mental processes of an individual (Pugalee, 2004). Verbal protocols are indeed valid as psychological data as much as other more typical sources of behavioural data obtained from experimentation (Ericsson & Simon, 1980).

In this study, validity was enhanced through referral of the verbal protocol and interview transcription to the participants for host verification. The protocols of the experts and novices were analyzed using the five research questions as a framework. This involved identifying the elements that were relevant to each of the research question. For the first, second, third and fourth questions, pre-determined categories of elements were imposed on the data, and the fit were observed. A coding scheme was developed and relative frequencies of salient events and strategies were obtained through the analysis of the verbal protocols. The fifth question was addressed after the mental model of the expert raters was identified and put into trial to help the novice raters.

4.2.1 Knowledge States Used by the Expert and Novice Raters in Assessing and Giving Feedback on Students' Writing

This sub-section presents the knowledge states used by both the expert and novice rater and later makes a comparison between them. The expert and novice raters are referred to as ER and NR respectively, according to the code names given to them in Appendix 13 to maintain anonymity of the participants.

(a) Knowledge States Used by Expert Raters

Writing a good English essay calls for observance to text conventions of the language. The **knowledge state** is a type of protocol representation (Newell & Simon, 1972) which identifies units of writing knowledge that were used by the expert and novice ESL teachers or raters in this study. This included their knowledge of clear writing, and evaluating and assessing skills needed in the task of rating and assessing writing. Table 4.1 shows the knowledge states used by the expert raters. It indicates that the expert raters had collectively used about 335 instances of knowledge states during the verbal protocol analysis of assessing task for both essays. The most widely used knowledge state was the choice of expression (94 times), which was followed by the choice of words (50 times) and clarity (32 times).

The expert raters had also referred to the use of tenses (27 times), preposition (15 times), development of paragraph (14 times), spelling (14 times), grammaticality (11 times), articles (10 times), organisation (9 times) and verb form (9 times) during the verbal protocol analysis task. The rest of the knowledge states such as determiners, subject-verb agreement, non-English terms, coherence, dangling sentences, sentence structures, run-on sentences, exemplification, capitalisation, thesis statement, adjectives, and punctuation were referred to between 2 to 6 times by the expert raters during the verbal protocol analysis tasks. Apart from that, pronouns and contraction were the least used or stated (only once).

Knowledge states	ER		Total Freq.	NR		Total Freq.
	Essay 1	Essay 2		Essay 1	Essay 2	
Grammar:						
Tenses	10	17	27	18	31	49
Prepositions	4	11	15	5	14	19
Grammaticality	8	3	11	2	3	5
Articles	9	1	10	11	4	15
Verb Form	6	3	9	1	1	2
Determiners	5	1	6	0	0	0
Subject verb agreement	0	6	6	6	3	9
Dangling Sentences	0	5	5	7	3	10
Sentence structures	4	0	4	5	0	5
Run-on sentences	1	3	4	3	1	4
Adjectives	2	0	2	2	0	2
Nouns	0	0	0	1	0	1
Pronouns	0	1	1	4	2	6
Contractions	0	1	1	2	2	4
Superlatives	0	0	0	1	0	1
Sentence variety	0	0	0	4	0	4
Infinitives	0	0	0	0	4	4
Mechanical aspects:						
Spelling	4	10	14	10	10	20
Punctuation	1	1	2	4	1	5
Capitalisation	0	2	2	0	0	0
Contents:						
Choice of expressions	50	44	94	8	1	9
Clarity	11	21	32	7	6	13
Non-English terms' equivalents	2	4	6	5	8	13
Exemplification	3	0	3	0	2	2
Development of ideas	0	0	0	4	0	4
Colloquial	0	0	0	1	0	1
Organisation:						
Development of paragraphs	12	2	14	20	13	33
Organization-related	5	4	9	4	0	4
Cohesion	0	0	0	0	1	1
Coherence	3	3	6	9	5	14
Thesis statement	0	2	2	1	1	2
Discourse markers	0	0	0	1	0	1
Vocabulary:						
Choice of words	18	32	50	9	10	43
Total No.	158	177	335	175	130	305

Table 4.1 Knowledge states used by the expert and novice raters

The knowledge states used by each of the expert raters are presented in the following sub-sections with reference to Appendix 13.

(i) ER1

Based on Appendix 13, ER1 referred to about 15 knowledge states while assessing sample essay 1, which is equivalent to 50 times in terms of frequency of use. Some of the main knowledge states he referred to during the verbal protocol analysis of assessing essay 1 were choice of expression, articles, determiners, clarity, choice of words, grammaticality, and sentence structure. Other than that, knowledge states like development of paragraph, tenses, form of verbs, spelling, punctuation, organisation, exemplification and adjective were referred to less frequently.

As for sample essay 2, ER1 had referred to about 13 knowledge states during the verbal protocol analysis, which is equivalent to 51 times in terms of frequency of use. Some of the main knowledge states he referred to during the verbal protocol analysis of assessing essay 2 were choice of expression and words, spelling, clarity, and dangling sentence tenses. Apart from that, knowledge states like preposition, capitalisation, contraction, grammaticality, non-English terms, organisation, and subject verb agreement were less frequently referred to.

(ii) ER 2

As clearly shown in Appendix 13, ER2 had referred to about 10 knowledge states while assessing sample essay 1, which is equal to 42 times in terms of frequency of use. The main knowledge states she referred to during the verbal protocol analysis of assessing essay 1 were choice of word, choice of expression, clarity, tenses and verb form. On the other hand, knowledge states such as spelling, adjectives, articles, run-on sentence and prepositions were referred to less frequently. In assessing sample essay 2, ER2 had referred to about 14 knowledge states during the verbal protocol analysis, which is equivalent to 60 times in terms of frequency of use.

Some of the major knowledge states she referred to during the verbal protocol analysis of assessing essay 2 were choice of word, choice of expression, run-on sentence, prepositions, subject verb agreement and tenses. Apart from that, knowledge states such as organization-related, coherence, clarity, capitalization, punctuation, articles, dangling sentence and verb form were less frequently referred to.

(iii) ER 3

As presented in Appendix 13, ER3 had referred to about 8 knowledge states while assessing sample essay 1, which is equivalent to 22 times in terms of frequency of use. Some of the main knowledge states he referred to during the verbal protocol analysis of assessing essay 1 were choice of word, coherence, choice of expression and development of paragraph. On the other hand, knowledge states such as organization-related, exemplifications, non-English terms' equivalents and grammaticality were seldom referred to.

As for sample essay 2, ER3 had referred to about 10 knowledge states during the verbal protocol analysis, which is equivalent to 19 times in terms of frequency of use. Some of the major knowledge states he referred to during the verbal protocol analysis of assessing essay 2 were choice of word, choice of expression and non-English terms' equivalents. Apart from that, knowledge states such as coherence, development of paragraph, organization-related, clarity, determiners, prepositions and grammaticality were less frequently referred to.

(iv) ER 4

Based on the Appendix 13, ER4 had referred to about 9 knowledge states while assessing sample essay 1 which is equivalent to 34 times in terms of frequency of use.

Some of the main knowledge states he referred to during the verbal protocol analysis of assessing essay 1 were choice of word, development of paragraph, organization-related, choice of expression, tense and grammaticality. On the other hand, knowledge states such as non-English terms' equivalents, spelling, prepositions were referred to less frequently.

As for sample essay 2, ER4 had referred to about 7 knowledge states during the verbal protocol analysis, which is equivalent to 47 times in terms of frequency of use. Some of the major knowledge states he referred to during the verbal protocol analysis of assessing essay 2 were choice of word, choice of expression and clarity. Apart from that, knowledge states such as spelling, pronoun, verb forms and thesis statement were less frequently referred to.

(b) Knowledge States Used by Novice Raters

Table 4.1 indicates that the novice raters had collectively used about 305 instances of knowledge states during the verbal protocol analysis of writing assessing task for both essays. The most frequently used knowledge state was tenses (49 times), which was then followed by the choice of words (43 times) and development of paragraphs (33 times).

Besides that, the novice raters had also referred to the use of spelling (20 times), prepositions (19 times), articles (15 times), coherence (14 times), clarity and non-English terms' equivalents (both 13 times), dangling sentence (10 times), choice of expression and subject verb agreement (both 9 times) during performance of the verbal protocol analysis. The remaining of the knowledge states such as thesis statement, organization-related, exemplification, development of ideas, punctuation, adjectives, infinitives, grammaticality, contractions, run-on sentences, sentence

structures, sentence variety, pronouns and verb forms were referred to between 2 to 5 times by the novice raters during performance of verbal protocol analysis. Apart from that, the least frequently used knowledge states were discourse markers, cohesion, colloquial, nouns and superlatives (only once).

The knowledge states used by each of the novice raters are presented in the following sections, with reference to Appendix 13.

(i) NR 1

Based on the Appendix 13, NR1 had referred to about 13 knowledge states while assessing sample essay 1 which is equivalent to 42 times in terms of frequency of use. Some of the main knowledge states he referred to during the verbal protocol analysis of assessing essay 1 were choice of words, development of paragraphs, spelling, articles and tenses. On the other hand, knowledge states such as organization-related, coherence, choice of expression, sentence structures, adjectives, prepositions, pronouns and subject verb agreement were referred to less frequently.

As for sample essay 2, NR1 had referred to about 12 knowledge states during the verbal protocol analysis, which is equivalent to 36 times in terms of frequency of use. Some of the major knowledge states he referred to during the verbal protocol analysis of assessing essay 2 were clarity, development of paragraphs, spelling, prepositions and tenses. Apart from that, knowledge states such as choice of words, coherence, thesis statement, non-English terms' equivalents, dangling sentence, articles and grammaticality were less frequently referred to.

(ii) NR 2

Based on the Appendix 13, NR2 had referred to about 21 knowledge states while assessing sample essay 1 which is equivalent to 45 times in terms of frequency of use. Some of the main knowledge states he referred to during the verbal protocol analysis of assessing essay 1 were choice of words, development of paragraphs, non-English terms' equivalents, spelling, punctuation, articles and tenses. On the other hand, knowledge states such as coherence, discourse markers, thesis statement, organization-related, clarity, development of ideas, choice of expression, sentence variety, contractions, sentence structures, grammaticality, nouns, prepositions and pronouns were referred to less frequently.

As for sample essay 2, NR2 had referred to about 15 knowledge states during the verbal protocol analysis, which is equivalent to 37 times in terms of frequency of use. Some of the major knowledge states he referred to during the verbal protocol analysis of assessing essay 2 were non-English terms' equivalents, prepositions, and tenses. Apart from that, knowledge states such as choice of words, coherence, clarity, exemplification, spelling, punctuation, contractions, articles, infinitives, grammaticality, pronouns and subject verb agreement were less frequently referred to.

(iii) NR 3

Based on the Appendix 13, NR3 had referred to about 17 knowledge states while assessing sample essay 1 which is equivalent to 33 times in terms of frequency of use. Some of the main knowledge states he referred to during the verbal protocol analysis of assessing essay 1 were choice of words, dangling sentence and subject verb agreement. On the other hand, knowledge states such as coherence, development of paragraphs, organization-related, clarity, non-English terms' equivalents,

development of ideas, sentence variety, punctuation, spelling, grammaticality, adjectives, pronouns, verb forms and tenses were referred to less frequently.

As for sample essay 2, NR3 had referred to about merely 5 knowledge states during the verbal protocol analysis, which is equivalent to 11 times in terms of frequency of use. Some of the major knowledge states he referred to during the verbal protocol analysis of assessing essay 2 were spelling and tenses. Apart from that, knowledge states such as choice of words, cohesion and non-English terms' equivalents were less frequently referred to.

(iv) NR 4

Based on the Appendix 13, NR4 had referred to about 16 knowledge states while assessing sample essay 1 which is equivalent to 55 times in terms of frequency of use. Some of the main knowledge states he referred to during the verbal protocol analysis of assessing essay 1 were choice of words, coherence, development of paragraphs, choice of expression, clarity, dangling sentence, run-on sentences and tenses. On the other hand, knowledge states such as organization-related, colloquial, development of ideas, articles, sentence structures, prepositions, subject verb agreement and superlatives were referred to less frequently.

As for sample essay 2, NR4 had referred to about 16 knowledge states during the verbal protocol analysis, which is equivalent to 46 times in terms of frequency of use. Some of the major knowledge states he referred to during the verbal protocol analysis of assessing essay 2 were choice of words, development of paragraphs, prepositions and tenses. Apart from that, knowledge states such as coherence, choice of expression, clarity, spelling, articles, contractions, dangling sentence, grammaticality,

infinitives, run-on sentences, subject verb agreement and verb forms were less frequently referred to.

(c) Comparison of the Knowledge States Used by the Expert and Novice Raters

Based on Table 4.1 and the description of findings presented earlier, it is apparent that expert raters managed to identify more instances of knowledge states than the novice raters. A total of 335 instances of knowledge states were given by the experts while the novice raters identified 305. This shows that the expert raters, given the experience they had, could be more meticulous in identifying language features of an essay.

In addition to that, a close scrutiny of the findings show that the novice raters tended to identify more grammatical-based knowledge states with little attention paid towards the content aspect. This is clearly shown in the results presented in Table 4.1 in which knowledge states related to the content of the essay such as development of ideas, choice of expressions and exemplification were not widely expressed by the novice raters as compared to their expert counterparts. Such situation depicts the nature of novice raters' assessing "method", which relied heavily on grammatical features. Nonetheless, both the novice and expert raters were rather balanced in terms of identifying knowledge states related to vocabulary or choice of words.

4.2.2 Conceptual Operators Used by Both the Expert and Novice Raters in Assessing and Giving Feedback on Students' Writing

This sub-section discusses the conceptual operators used by both the expert and novice raters and later makes a comparison between them. The **conceptual operation** in this study refers to an inferred cognitive process which modifies (adds, eliminates) existing or currently active knowledge states and produces new, active knowledge

states (Hassebrock & Prietula, 1992). In response to the assessing writing task which is also a specific data cue, a given segment of verbal protocol by the raters would constitute one or more knowledge states and a conceptual operation that produces the associated knowledge state or states.

Conceptual operations	Conceptual operators	ER		Total freq.	NR		Total freq.
		Essay 1	Essay 2		Essay 1	Essay 2	
Data examination	Read	0	0	0	3	1	4
	Identify	22	48	70	40	38	78
	Examine	6	2	8	32	4	36
Data exploration	Examine	0	0	0	1	0	1
	Scan	1	1	2	1	0	1
	Search	5	0	5	0	0	0
	Elaborate	2	10	12	2	3	5
	Note absent data	4	1	5	0	0	0
Data explanation	Infer	30	38	68	25	21	46
Hypothesis-generation	Trigger	9	1	10	5	2	7
	Further-specification	0	0	0	2	0	2
	Association	0	0	0	0	0	0
	Generalization	0	0	0	2	1	3
Hypothesis-evaluation	Confirmation	7	2	9	0	3	3
	Disconfirmation	0	0	0	0	0	0
	Discrimination	0	0	0	0	0	0
	Casual relationship	0	0	0	1	0	1
Discrepancy-processing	Recognition	8	3	11	5	4	9
	Resolution-explain away	5	3	8	4	5	9
	Ignore	0	0	0	2	0	2
Meta-reasoning	Experiential-memory	10	13	23	14	9	23
	Self-evaluation	80	73	153	40	14	54
Sumarisation	Repeat-data	1	3	4	4	7	11
	Repeat hypothesis	0	1	1	0	0	0
			Total no.	389		Total no.	295

Table 4.2 Conceptual operators used by expert raters and novice raters

Based on Table 4.2, the main conceptual operations used by both the expert and novice raters during the verbal protocol analysis were data examination, data exploration, data explanation, hypothesis generation, hypothesis evaluation, discrepancy processing, meta-reasoning and summarization. They were used to characterize distinct segments of a raters' problem-solving behaviour. The emphasis given to the use of these conceptual operations during the verbal protocol analysis

made the difference between the expert and novice raters in the problem-solving process of assessing writing. To give a more detailed representation of knowledge and reasoning behaviour required by the task, each basic conceptual operation was further analyzed and interpreted into raters' lines of reasoning.

(a) Conceptual Operators Used by Expert Raters

Based on the Table 4.2, the expert raters had referred to about 15 conceptual operators while assessing the two sample essays which is equivalent to 389 times in terms of frequency of use. Some of the main conceptual operators they referred to in the verbal protocol analysis of assessing both essays were self-evaluation (meta-reasoning), identify (data examination), infer (data explanation) experiential-memory (meta-reasoning), elaborate (data exploration) and recognition discrepancy-processing). On the other hand, conceptual operators such as trigger (hypothesis generation), confirmation (hypothesis-evaluation), examine (data examination) resolution-explain away (discrepancy-processing), search (data exploration), note absent data (data exploration) repeat data (summarisation), scan (data exploration) and repeat hypothesis (summarisation) were referred to less frequently.

The conceptual operators used by each of the expert raters are presented in the following sections with reference to Appendix 14.

(i) ER1

Based on Appendix 14, ER1 recorded 69 occurrences of conceptual operators in essay 1 and majority of those operators fall under meta-reasoning, namely experiential-memory (10 occurrences) and self-evaluation (22 occurrences). This is followed by

infer (18 occurrences). Other operators used were examine (data examination), search (data exploration) and recognition (discrepancy-processing).

In essay 2, ER1 indicated 54 occurrences of conceptual operators with identify (data examination) recorded the highest occurrences (24). Infer (data explanation) remained high with 15 occurrences. Lesser self-evaluation (meta-reasoning) is noted in essay 2 as compared to essay 1. There was also dependency on experiential memory, recognition, note absent data and examine (data examination).

(ii) ER 2

As shown in Appendix 14, a total of 49 occurrences of conceptual operators were recorded by ER2 when assessing essay 1. Self-evaluation (meta-reasoning) was referred to the most, followed by identify (data examination). Other notable operators used in essay 1 were elaborate (data exploration), note absent data (data exploration) and infer (data explanation). However, in essay 2, there is a significant increase in the frequency of conceptual operators. Out of the 69 occurrences of conceptual operators, 30 occurrences were noted for self-evaluation (meta-reasoning), 17 occurrences for infer (data explanation) and 14 occurrences for identify (data examination). None of the operators for data exploration, hypothesis generation and hypothesis evaluation was used by ER2 in essay 2.

(iii) ER 3

ER3 referred to the least conceptual operators as compared to the other four expert raters as shown in Appendix 14. 28 occurrences of conceptual operators were noted in essay 1 with self-evaluation (meta-reasoning) being the highest, followed by infer (data explanation) and identify (data examination).

In essay 2, the conceptual operators were only referred to 24 times, with self-evaluation (meta-reasoning) being referred to the most. Infer (data explanation), identify (data examination) and experiential memory (meta-reasoning) were also referred to during the process of assessing essay 2.

(iv) ER 4

ER4 referred to eight conceptual operators when assessing essay 1, with a total occurrences of 44 (refer to Appendix 14). Of the total, self-evaluation (meta-reasoning) recorded 19 occurrences, followed by trigger (hypothesis-generation) and confirmation (hypothesis-evaluation). On the other hand, in essay 2, ER4 showed higher reference to self-evaluation (meta-reasoning) with 28 occurrences. The other conceptual operators referred to are elaborate (data exploration), identify (data examination), infer (data explanation) and repeat-data (summarization).

(b) Conceptual Operators Used by Novice Raters

Based on the Table 4.2, the novice raters had referred to about 18 conceptual operators while assessing the two sample essays which is equivalent to 295 times in terms of frequency of use. Some of the main conceptual operators they referred to in the verbal protocol analysis of assessing both essay were identify (data examination), self-evaluation (meta-reasoning), infer (data explanation), examine (data examination), experiential-memory (meta-reasoning), and repeat data (summarisation).

On the other hand, conceptual operators such as recognition (discrepancy-processing), resolution-explain away (discrepancy-processing), elaborate (data exploration), read

(data examination), generalization (hypothesis-generation), confirmation (hypothesis evaluation), further-specification (hypothesis-generation), ignore (discrepancy-processing), examine (data exploration), scan (data exploration) and causal relationship (hypothesis-evaluation) were referred to less frequently.

The conceptual operators used by each of the novice raters are presented in the following sections with reference to Appendix 14.

(i) NR1

In essay 1, NR1 referred to 12 types of conceptual operations. The dominant ones are infer (data elaboration), identify (data examination), and self-evaluation (meta-reasoning) as indicated in the number of occurrences shown in Appendix 14. There was minimal use of hypothesis-evaluation and hypothesis-generation.

Surprisingly, in essay 2, NR1 showed less reference to conceptual operations with only 15 occurrences recorded. Of the total, repeat-data (summarization) and infer (data exploration) showed prominent frequency. There was no indication of data exploration at all.

(ii) NR 2

In essay 1, NR 2 referred to 15 types of conceptual operators with a frequency of 46. Identify (data examination) showed highest occurrences, followed by examine (data examination), infer (data explanation) and self-evaluation (meta-reasoning). The occurrences of discrepancy-processing were also higher in NR2 as compared to other novice raters.

In contrast, in essay 2, the frequency of conceptual operators by NR2 was slightly lower with a total of 40 occurrences. 21 occurrences were noted for identify (data examination), followed by infer (data explanation). It seems that in essay 2, NR2 depended more on data examination in assessing.

(iii) NR 3

As shown in Appendix 14, NR3 referred to eight conceptual operations with total frequency of 31. NR3 referred to identify (data examination) the most with 12 occurrences as well as self-evaluation (meta-reasoning) with 8 occurrences. Other conceptual operators used by NR3 are infer (data explanation), recognition (discrepancy-processing) and resolution-explain away (discrepancy-processing).

As for essay 2, NR3 referred to the least conceptual operations in comparison with other novice raters. Only 5 types of operations are noted, with a mere 12 occurrences. Of this total, self-evaluation (meta-reasoning) was predominantly used, followed by repeat-data (summarization), and experiential-memory (meta-reasoning). It is interesting to note that NR3 used none of operators within data examination and data exploration when assessing essay 2.

(iv) NR 4

NR4 recorded 53 occurrences of conceptual operators in essay 1, as indicated in Appendix 14. Examine (data examination) was the most referred to with 24 occurrences. Self-evaluation (meta-reasoning) was also quite high with 15 occurrences. NR4 also referred to identify (data examination), elaborate (data exploration), infer (data exploration), trigger (hypothesis-generation) and confirmation (hypothesis-evaluation).

In essay 2, NR4 referred to eight types of conceptual operations with identify (data examination) remained primarily high (17 occurrences). This is followed by infer (data explanation), self-evaluation (meta-reasoning) examine (data examination), resolution-explain away (discrepancy-processing).

(c) Comparison of the Conceptual Operators Used by the Expert and Novice Raters

As noted earlier, both the expert and novice raters had used 8 conceptual operations in their VPA. They were data examination, data exploration, data explanation, hypothesis generation, hypothesis evaluation, discrepancy processing, meta-reasoning and summarization.

Based on Table 4.2, the expert raters had only used 15 conceptual operators as compared to the novice raters who used 18 conceptual operators. However, in terms of frequency of use, the expert raters used more instances of conceptual operators (389 times) as compared to the novice raters (295 times). Common conceptual operators used by both expert and novice raters were identify and examine (data examination), scan and elaborate (data exploration), infer (data explanation), trigger (hypothesis-generation), confirmation (hypothesis-evaluation), recognition and resolution-explain away (discrepancy-processing), experiential-memory and self-evaluation (meta-reasoning), and repeat-data (summarization), although frequency of use differed.

Conceptual operators like search and note absent data (data exploration), and repeat hypothesis (summarization) were seen to be used only by the expert rater but usage is quite negligible. On the other hand, conceptual o perators like read (data

examination), examine (data exploration), further-specification and generalization (hypothesis-generation), causal relationship (hypothesis-evaluation), and ignore (discrepancy-processing) were seen to be used only by the novice raters.

Meta-reasoning in this study was used by the raters to evaluate the reasoning process during the VPA of the given tasks. There are two conceptual operators that are involved in this operation, experiential-memory and self evaluation. The raters used self-evaluation to criticize or critique the adequacy of specific diagnostics hypotheses or general diagnostic conclusions in the students' writing. On the other hand, they used experiential-memory to recall related information and also retrieve related general knowledge. Based on Table 4.2, it is obvious that the expert raters had given a lot of emphasis on self-evaluation which was referred to 153 times as compared to the novice raters who referred to it for only 54 times. However, coincidently both the expert and novice raters had the same number of usage for experiential-memory (23 times).

Another significant observation from Table 4.2 was the emphasis given by the expert raters on the conceptual operator - infer (data explanation) which was very much content and meaning focus. Here, basically the rater's goal was to interpret the significance or role of a given data cue and the operator - infer was used to explain the developmental course of a particular cue with respect to a good piece of writing.

Unlike the expert raters, the novice raters seemed to focus more on surfaced errors when they concentrated on using the conceptual operator - identify during the data examination process. During this process, the raters merely selected and examined cues and made initial interpretations of the finding in the student's writing. Through

the conceptual operator - identify, the raters would selectively identify or repeat a specific cue or cues from that particular set of data.

4.2.3 Differences in Lines of Reasoning between Expert and Novice Raters in Assessing and Giving Feedback on Students' Writing.

This section discusses the differences in the lines of reasoning between the expert and the novice raters. Specifically, the data from the quantified knowledge states and conceptual operators were analysed in relation to the interview data to distinguish the lines of reasoning between the two groups of raters. Later, the results would be used to construct the conceptual mental models of both the expert and novice raters.

(a) The Expert Raters' Lines of Reasoning

(i) ER 1

Based on table 4.1, of the total of 15 knowledge states referred to by ER 1, the two most widely used knowledge states that expert raters employed in assessing both sample essays, were the choice of expressions and words. In essay 1, ER 1 referred to choice of expressions 19 times, the highest among the expert raters. It seems to indicate that the words used as well as the choice of expressions are the decisive factor for ER 1 to distinguish which essay is better. In the exchange below, ER 1 provided the following related information:

> Example 1
>
> …a B (essay) to me would mean that the command of the language is there but the "sophistication" of it is not there. So sophistication…I would put it as complex structures and vocabulary that is precise.
>
> (ER1/VP/Transcript/13.6.07)

As an expert rater, ER 1 was very thorough in his assessment. When embarking on an assessing task, his first main priority was to get an overall glimpse of what the writing was all about. Based on the data from the conceptual operators used, it is clear that ER 1 relied a lot on data explanation (infer) as well as data examination in trying to figure

out which essay deserves a better grade. Infer recorded a frequency of 33 occurrences, the highest among all the raters. In particular, he provided the following response in Example 2 to the researcher's question about what would be his initial step in assessing a piece of writing.

Example 2

> Normally, if you give me some essays to look at, let say three…I'll read …definitely the first paragraph first. That's the first thing I will do. I don't need to read the rest of the paragraphs…I just need to read one. So, when I read it, basically…what I'm looking for…like I said earlier…the ability of the candidate to communicate to me… From there, I would be able to judge the level of competency the candidate has.
>
> (ER1/VP/Transcript/13.6.07)

ER 1 seemed to be able to conceptualise a well-written essay and differentiate it from the less 'competent' ones. For instance, in Example 3, he clearly showed that he knew how to differentiate between a competent writing from an excellent one.

Example 3

> Basically, when we say he is C, we're saying that he is competent… competent as I said, he has mastered the basics. That's what we'll look for. Basics here…you can check out primary school materials…primary six materials. If he can master primary six materials, that's good enough. We can say he's competent…let's not put competency and excellence…they are two different things. Competency means the basics, excellence means that's an A. When we say basic competency, we're looking at primary six materials. That's what I'll be looking for, because primary six, tenses is clean, there are no errors basically, structures are fine…if you're asking for more than that, you're actually asking for excellence…that's a different story.
>
> (ER1/VP/Transcript/13.6.07)

(ii) ER 2

ER 2 stressed a lot on the choice of expression in both essays. Similar to ER1, ER 2 recorded a very high frequency of references to choice of expressions (32 occurrences). During the interview, when asked to comment on how she would usually assess her students' writing, ER 2 divided her comments into three categories: structure, content and language. Within these three categories, choice of expressions

shapes the quality rating of the essays given. This can be seen from the interview with ER 2 in Example 4.

Example 4

> Actually, when I mark an essay, I will look at the criteria. Generally I will look at three perspectives. One is structure, the other one is content and the other one is the language itself.
>
> ER2/Interview/Comments/14.6.07

She indicated in her interview response that for structure, she would be looking for paragraphing. This is clearly stated as given in Example 5:

Example 5

> Structure…I'll be looking at…whether there is paragraphing and then… is there any intro, body and conclusion. I'll be looking at something like this.
>
> ER2/Interview/Comments/14.6.07

As for organisation, she would be looking at the main points (topic sentences) and how they develop into paragraphs, which would give her the impression of how mature students were as clearly shown in Example 6:

Example 6

> I will be looking at organisation, the details…how good the details…how good are the details. Okay, so we are looking at main point …elaboration...the organisation of ideas itself. How mature the students are…and so on.
>
> ER2/Interview/Comments/14.6.07

After deciding on whether meaning (particularly in the correct and appropriate choice of words) was obvious in the writing or not, she would then decide on the grade as clearly articulated in Example 7 and Example 8:

Example 7

> …band D essay...there will be patches here and there… I would ask myself let's say there are patches…I'll be asking the question like does it hamper reading…because of band E. (If it) hamper reading, we go to band E, so if not or quite, I'll adjust the marks from there.
>
> ER2/Interview/Comments/14.6.07

Example 8

> For me, this particular one (one of the example essay), is a cross between C and D. The story line is quite cute I would say…actually at first, just look at the story line. I would want to give a C because I think the writer is at least very ambitious…he tries his best…he wants to be humorous. But unfortunately, it's not secured at the end. And then sentence structure that he used…I think for a form five kid, I think it's very good. So I give a high D…I would not give a C because of the mistakes…actually if I'm "cruel" enough, I can go further down. But for the work he has done so far it's around 30. I think I would give D30…because there are some parts that really make me think somehow.
>
> ER2/Interview/Comments/14.6.07

ER 2 relied the most on meta-reasoning (self-evaluation) with a total occurrence of 53, as compared to other expert raters. This is reflected in the interview data earlier which revealed her systematic way of reasoning when it comes to evaluating an essay. ER 2 also relied on data explanation (infer) and data examination during the process of evaluating the essay. A lot of inferences are done in determining the meaning of the sample essays.

(iii) ER 3

When doing the 'thinking aloud' task assigned to him, ER 3 did not really go into details on how he would assess students' writing. Obviously, what he generally did was skimming through the writing to determine the general idea of what it was all about before awarding marks based on holistic/impression grading. This is apparent in the recorded frequencies of knowledge states and conceptual operators. ER 3 used the least of both but revealed strong preference in focusing on the choice of expressions and words (15 of the 44 recorded knowledge states). He also focused his assessment based on the students' command of grammar, sentence structure and vocabulary as he went through the pieces of writing. These things were being highlighted by him as in the Example 9:

Example 9

> Generally in marking an essay question, I will first skim through the essay because I don't want to find myself penalising the kids…the students for the grammatical (errors), the vocabulary and the error...the semantics…so what I'll do is I'll quickly read through to see whether I understand what the candidate is writing about and where I would put them in the grade….between A, B, C, D…having get that impression I would then read the essay and see whether I can give the student a better grade from the general grade by looking at the grammar, sentence structure and err...vocab.
>
> ER3/Interview/Comments/14.6.06

ER 3 was very particular in the kind of positive feedback his would give to students. As shown in Example 10, he would highlight strengths and weaknesses in the sentence structures or the vocabulary used so that students would know how to improve their writing.

Example 10

> …if I were to give them feedback…although the grades…the poorer one was graded as E and the other one only a weak C…I would encourage them. The thing I would highlight to them would be…they were able to get the meaning across…and…and I would also highlight the sentences that they did very well. I will also highlight for the second candidate the vocabulary that she has picked up obviously through a lot of reading or maybe listening to the teacher in class. So somehow she is able to get all that. And I would highlight (them) … then I would call them to carry on with that…
>
> ER3/Interview/Comments/14.6.06

As said earlier, ER 3 did not verbalise his thought much during the VPA task. However, he knew exactly what to look for when assessing his students' writing. He even illustrated how the markers' impression of sentence can be changed from a C grade to a B or an A grade. ER 3's data for conceptual operators also revealed the same trend as there is a high dependency in using meta-reasoning (self-evaluation) and data explanation (infer) in the rating process. In a response to how he would help his students to improve their writing, he provided the following information in Example 11:

Example 11

> I would also highlight sentences that were not so good in class…write out on the board or in papers and hand it back to them…and probably get the class… say…how can you make these (an) A-grade sentences. Okay? This sentence is now graded as C because of some errors like the one that starts with (an) error in preposition. I would write the whole sentence down in class and I would tell them…this one is a C sentence right? It… gives an impression of a C essay or a D essay. If you were to change some mistakes…if you were to address some of the mistakes then you could change the impression from a C to probably B or an A. I will show them how easy it is to change impression or the marker's (assessors) impressions…
>
> ER3/Interview/Comments/14.6.06

(iv) ER 4

ER 4 produced many more comments as compared to the other expert raters. Not only he had near native English proficiency as an ESL writing assessor, the greater number of comments he produced may had come from his teaching experience and also as a chief marker for English Paper II in SPM examination. His data for knowledge states revealed a similar trend with other raters in terms of high frequencies of choice of expressions and words. However in essay 2, he mentioned clarity rather often as compared to the rest of the raters. In terms of conceptual operators, ER4 also depended frequently on meta-reasoning especially self-evaluation. During the verbal protocol analysis task, he used codes to indicate the types of errors in students' writing. These codes are similar to the one used in marking SPM English Paper II. He also identified errors by underlining, crossing out and writing comments in above the text and in the margin. When given the task of assessing writing, ER 4 showed that he placed his priority on meaning and language. This is pretty obvious from his interview response in Example 12 when asked on how he would mark a piece of writing.

Example 12

> When I look at essays, I'm actually looking at basically at two things. The first is I'm looking at for the ability to communicate their ideas to us…the second is command of the language. I'll be looking for these two things. What do I actually look for when I say the ability to communicate…I'm actually looking at whether or not they can convey their ideas to me in English. Of course, they can come in many forms...but what I need to know is that when I read the essay, is it standard English? Number two…when I read it, do I have to translate it from another language, is there a need to paraphrase the thing or I can say that it's not something I would recognise but I would need to do a bit of work. So, I'm looking for what I would call…I'll use the phrase as in it comes clearly to me without any element of doubt. I do not want to see things that I need to infer or make a decision. So that's the competitive aspect I'm looking for.
>
> ER3/Interview/Comments/14.6.06

Overall, ER 4 showed concern over language use, content, and development of the essays. Although he would look for the flow of ideas when he read the essays, he might catch a lot of grammatical errors. From the observation during the verbal protocol analysis, he seemed to be able to process information on both form and content simultaneously. He seemed to be very expressive in his comments on aspects such as the students' ability to communicate their ideas and the command of the language, including sentence structures, vocabulary, paragraphing skill and so on. He articulated all these in the interview response as in Example 13 and Example 14:

Example 13

> The other aspect that I'm looking for is language aspect. I'm looking at the command of the language…meaning. Of course, I'm looking at sentence structure, I'm looking to see whether or not the command of it…meaning here I'm looking at whether or not they are good at it...good meaning for example simple, compound, complex structures…are they able to make that…use it correctly and how frequently…meaning is it correct all the time? When I look at the sentences, that's what I would think… I'm looking for better structure. Also, when I look at the structure, it is so badly written that I have to restructure it every time via language. When I'm looking at vocabulary, I'm looking at the word…in the sense that does it convey the precise meaning I want. So, that's what I'll be looking at.
>
> ER3/Interview/Comments/14.6.06

Example 14

> …I'm looking at his stability to give me words that is precise enough. And then at the same time, I'm looking at words that would indicate that his vocabulary is wide enough ….wide enough means it's like the word run…I would expect, does he knows other words beside run? Maybe jog, maybe stroll…meaning I'm looking for his other choices, other alternatives…is he going to give me that?
>
> ER3/Interview/Comments/14.6.06

Although most of his comments remained at content level, he also gave feedback on organisation and surface level (i.e. spelling and grammaticality), as defined in Section 3.8.1(i). This is as shown in the interview response in Example 15.

Example 15

> …looking at paragraphing…can he arrange it logically, paragraphing and more than that. Of course, I'll be looking for the little aspects like adverbs, does he get it correct? This would be at the higher level. So when we talk about grammar…of course I'm looking at tenses, articles, these are little little small things I look out for…because these sort of things tell me his command of language. Basically I'll put it like this…is he competent, when I say he's competent meaning that I would expect the basic elements that what he's supposed to have learn in primary school…
>
> ER3/Interview/Comments/14.6.06

(b) The Novice Raters' Lines of Reasoning

(i) NR 1

As a beginning teacher who had just graduated, NR 1 had no problem skimming through each piece of writing given to him during the VPA task to decide if it was a good piece of writing. After skimming through, only then he would scanned through the paragraphs for clarity of ideas and identify errors. The knowledge states that he indicated are mainly choice of words and expression as well as articles. Based on what he said in Example 16, though he would try to identify all the errors in that particular piece of writing but the focus was more on the meaning. He used this approach to decide if errors committed had impeded meaning which the writer intended to convey.

Example 16

> Normally I would just read through once and get the tips but sometimes err, if it's a good piece of writing, meaning err…. Good in the sense of I can understand everything the writer is trying to say without really struggling. Then I would go paragraph by paragraph err… because I know I can… I can get the meanings straight away without having to read so many times. So… I will read (a) paragraph, get the whole meaning of the-that paragraph, and I will identify whether there's grammar error(s) or spelling errors…
>
> NR1/Interview/Comments/15.6.06

Overall, he would focus on the clarity of ideas (meaning) first, then would proceed to aspects of grammar and spelling errors as clearly stated in Example 17:

Example 17

> My focus would be on the meaning, whether I get the meaning first and then I will look through paragraph by paragraph. I'll mark all the errors; just tick or just make a comment or whatever and then… after I finish one paragraph, I will try to look back at the errors that I have marked whether it really is an error that will really disturb the meaning or it can be accepted…
>
> NR1/Interview/Comments/15.6.06

When dealing with less competent writing, NR 1 said that he would really have to read once and sometimes twice to get the meaning first of what the writer was trying to say before marking for the essay for errors. He stressed this strategy of marking towards the end of the interview by saying that when he was "marking an essay, the focus was on the meaning first, then the grammar and the spelling error" (NR1/Interview/ Comments/ 15.6.06). In terms of conceptual operators used, NR 1 showed high dependency on data explanation (infer) and data examination (identify) though there was a pertinent drop in the use of conceptual operators when evaluating essay 2. This is probably essay 2 has lesser problems than essay 1, while NR 1 was able to evaluate the essay on a surface level without going deep into distinguishing the errors of the language items used in essay 2.

(ii) NR 2

NR 2 provided comments mostly at sentence level errors. This is clear in the number of knowledge states mentioned by NR 2, as she referred to more types or language items than the rest. However, choice of words and expressions were still dominant. She underlined and wrote comments below the text, circled out, slashed out, and used codes such as N for noun and T for tense, to indicate the category of the errors. Apart from that, she also gave emphasis on the meaning aspect for every paragraph. In the exchange below, NR 2 provided the following information:

> Example 18
>
> Basically, when I mark students' writing I would first mark all the errors. At the same time I will try to see if the errors impede meaning. By identifying all the possible errors, it would help me decide if the essay conveys the appropriate meaning.
>
> NR2/Interview/Comments/15.6.06

She preferred clear and simple ideas expressed in students' writing. When she looked at students' writing, she searched for a main idea and how it was supported in the paragraph. She slashed out or underlined errors, and put ^ marks indicating that some words needed to be added to be correct. She also '…underlined the topic of each paragraph and/or asked for a transition signal appropriate for the paragraph' (NR2/VP/ Transcript/15.6.06). Finally, she wrote a grade without any end-note comment and she has '…highlighted all the errors throughout the essay' (NR2/Interview/Comments/ 15.6.06). This is supported by the data from the identified conceptual operators as she depended highly on data examination (identify) with 31 occurrences (the highest among the novice raters) and less reasoning process was involved during the rating process.

(iii) NR 3

NR 3 articulated that when marking essays, she preferred reading the good ones first. To begin with, she would skim through the essays for general idea. After that, she will scan through the paragraphs for errors as what she did during the verbal protocol analysis. She would slashed out or underlined errors, and put ^ marks indicating that some words needed to be added to make the sentence correct. This was also indicated in her interview response when asked how she would mark an essay as given in Example 19.

Example 19

> When I get a pile of paper to mark, normally, I'll choose the best students…I like to read the good ones first. When I get to the paper, what I'll do is I'll read the paper from beginning to the end first…then after that…I go from paragraph to paragraph. I'll look for whatever errors, whatever points I need to…whatever marks I need to give. Then…from there, after that…after I've chosen the best one, then I'll go to the medium one. The one that's not so good, in a way, I'll see how much that medium guy has improved after a period of learning.
>
> NR3/Interview/Comments/25.4.07

Based on the Example 19, the medium essays would be given the same treatment as the good essays. As reflected in Example 20, she opined that students who write good essays will always produce good essays but the medium ones will show the difference from their previous writing.

Example 20

> …the good ones will always be good throughout but the ones that's medium…(we) can easily see their improvement in their writing…after certain period of learning. So I like to do that…choose the one that is from the intermediate group and then compare to what he has done in the first test. I'll see how much he's improved. Then I'll mark his paper similarly…like that also.
>
> NR3/Interview/Comments/25.4.07

NR 3 would normally read the weak essays last adding that it was a much easier task. For her, marking the weak essays was simply by just underlining the obvious errors to

which she would later give face-to-face feedback. This response was indicated clearly in the interview discourse in Example 21:

Example 21

> …normally, I'll leave the weak one to the last because that's the easiest to mark. What I'll do is I'll just…the weak one normally have very short essays…very obvious mistakes…normally I'll leave it to the last. So, how was it…the last one…the the...the weak one would be err….there's nothing I can do, all I can do is to underline…underline…then I'll go and see them personally.
>
> NR3/Interview/Comments/25.4.07

Overall, she focused more on error identification to help her assess her students' writing. The data from VPA supported this as she recorded the highest frequency in data examination (identify). However, it is interesting to note that there were less processes involved in rating an essay for NR 3 because other conceptual operators recorder either zero or very low frequency. In total she recorded only 43 occurrences of conceptual operators, which is the least among the novice raters.

(iv) NR 4

As a beginning teacher, NR 4 seemed to be very confident in assessing the two pieces of writing during the VPA. He skimmed through first and last paragraph first to get an overall idea of what the essay was all about. He used the holistic approach of assessment to determine students' writing proficiency level as illustrated in the discourse as in Example 22:

Example 22

> Usually what I will do before I mark a test paper especially in writing is that I will read first paragraph and I will read the last paragraph. Because generally I think that if you read the first paragraph and the last paragraph, you get the whole idea of the direction of the story first and ah…secondly, by doing so…you are able to give a holistic view on student's level of the English language and more or less set the benchmark in my mind as to how good this student will be and how good I hope he will be actually.
>
> NR4/Interview/Comments/15.6.06

NR 4 also scanned through the writing to look for a storyline which would show how effective students had developed the ideas presented in the paragraphs. This was clearly stated by him as in Example 23:

Example 23

> … I would scan the whole story as I go along and ah…what I look for is story line. I think it is very important that students are able to develop their ideas.
>
> NR4/Interview/Comments/15.6.06

As the rest of the beginning teacher would do in assessing writing, he also focused a lot on errors identification like what he said in Example 24 as follows:

Example 24

> …I'm very particular when it comes to past tense; I feel that if you were to write a story, it is better to be in past tense. It's easier to just have everything in past tense…than a mixture of tenses.
>
> NR4/Interview/Comments/15.6.06

Based on the data from knowledge states, he recorded the highest number of occurrences with a total of 101 with high frequency for choice of words and expressions and tenses. As revealed in the interview data, he would be on the lookout for good use of expression by the students to decide if there was a variety of language use in the particular piece of writing, apart from identifying the presence of good ideas, good sense of humour and originality in it. The conceptual operators data revealed that he emphasised a lot on data examination (identify and examine). Such emphasis is obviously seen in the discourse in Example 25:

Example 25

> What I will do is that every time there is a good use of expression, for example, "she waited patiently", I get very impressed. My students who are able to write things like that show that they are able to use a variety of language rather than just simple and basic English. I will definitely give credit to those who have good ideas, good sense of humour even, and originality in their story. That's how I would mark.
>
> NR4/Interview/Comments/15.6.06

One commendable approach in his method of assessing writing was the way he was selective in marking students' writing. He would always mark the weak writings first to avoid setting high benchmark that may influence his decision making as illustrated in what he said in Example 26:

Example 26

> …I will usually start marking the weaker students first and then move on to better students…because if I were to read the better students first then the benchmark would definitely be higher and it would not be fair to the weaker students. I think in marking essays especially, you have to encourage them to write more. If you were to mark them very low…it would discourage them even more. So that is basically my strategy in marking essays.
>
> NR4/Interview/Comments/15.6.06

Overall, based on the Example 26 above, NR 4 emphasized on giving encouragement for students to write more, thus he would avoid giving them very low assessment.

(c) Comparison between the Expert Raters and Novice Raters' Lines of Reasoning

Knowledge states	ER		NR	
	Freq.	%	Freq.	%
Grammar	101	30.15	140	45.9
Mechanical aspects	18	5.37	25	8.20
Content	135	40.3	42	13.77
Organisation	31	9.25	55	18.03
Vocabulary	50	14.93	43	14.10
Total	335	100	305	100

Table 4.3 Comparison of knowledge states used between expert and novice raters

Based on the Table 4.3, the novice raters seemed to focus more on grammar as the most important concern when attending to students' writing. As compared to the expert raters, they attended to surface-level concerns like grammar and mechanical errors much more except for organisation where they were most concerned with the development of paragraphs. The novice raters highlighted grammatical errors, organisation and mechanical errors much more than the expert raters. The use of

content development and expressive qualities were not so emphasized on as the expert raters did.

Unlike the novice raters, Table 4.3 shows that the expert raters focused more on the content of the essay especially on the aspects of choice of expression and clarity as shown earlier in Table 4.1. For the expert raters, grammar was their second priority, followed by vocabulary, organisation, and finally mechanical errors. This pattern seemed to be directly attributable to the kind of assessing experience which the expert raters seemed to be accumulating over their years of classroom practice and marking examination paper at PMR and SPM level. The expert raters did not overlooked grammatical errors but they seemed to focus on errors that inhibit communication.

On the other hand, the novice raters may attend more to surface-level errors because they the easiest to detect and respond to. However, their comments on content may demand a higher degree of judgement and most likely take more time and so was attended to less frequently or in less detail. Leki (1991), in fact, speculates that because errors in grammar and mechanics are more concrete than meaning-related problems, they are relatively easier to correct.

Based on Table 4.4, it shows that the main conceptual operations used by both the expert and novice raters during the verbal protocol analysis were data examination, data exploration, data explanation, hypothesis generation, hypothesis evaluation, discrepancy processing, meta-reasoning and summarization. The emphasis given to the use of these conceptual operations during the verbal protocol analysis makes the difference between the expert and novice raters' line of reasoning in the problem-solving process of assessing writing. The knowledge content of the line of reasoning

serves to discriminate expertise as well as individual differences between the expert and novice raters in assessing writing.

Conceptual operations	ER		NR	
	Freq.	%	Freq.	%
Data examination	78	20.05	118	40
Data exploration	24	6.17	7	2.37
Data explanation	68	17.48	46	15.59
Hypothesis generation	10	2.57	12	4.07
Hypothesis evaluation	9	2.31	4	1.36
Discrepancy processing	19	4.88	20	6.78
Meta-reasoning	176	45.24	77	26.10
Summarization	5	1.29	11	3.73
Total	389	100	295	100

Table 4.4 Comparison of conceptual operations used between expert and novice raters

Based on Table 4.4 which displays the conceptual operations being deployed during the verbal protocol analysis, the expert raters placed meta-reasoning as top in their priority list while the novice raters focus more on data examination. This pattern was expected as the expert raters placed priority on the content of writing which employs high level cognitive process such as meta-reasoning to help in problem-solving. As compared to data examination, it was rather obvious that meta-reasoning may demand a higher degree of judgement that required the raters to fall back on their vast experience and exposure to marking and assessing writing. Obviously, the expert raters have an advantage on this matter as compared to the novice raters.

The expert raters also did more data exploration and data explanation as compared to the novice raters. This pattern complements the extensive use of meta-reasoning during the verbal protocol analysis. Thus, the expert raters were able to articulate more concrete comments on how they assessed writings during the verbal protocol analysis and be more effective in their decision making. The Sample 1 (Table 4.5) is the transcript fragment and coding of ER 4 on essay 1 which illustrates the lengthy in-depth comment and variety of conceptual operators being used by the expert rater.

Sample 1

The first sentence has a spelling error (written in Bahasa Melayu). There are also wrong word use where 'prepaid' is mistaken used for 'prepared', and 'live' is used instead of 'stay'. Initial assessment of the essay is likely to put it under band 'U' as one has to infer meaning on the ideas put across.

[(1) At the 27th Disember…okay, never mind, small matter – BM spelling. …2003. [(2) my family and I 'prepaid'…okay, problem with spelling again.] [(3) …prepared to (to go to) Damai Beach Resort… 'Prepaid'? What does he mean by 'prepaid'?] [(4)We live there for four day(s) three night(s).] [(5) I went there by the big car of my father.] [(6) Oh, he…they went to Damai Resort.] [(7) So 'prepaid' is not the right word.] [(8)Prepared to go?] [(9) I'm not sure what he wants…'prepaid'.] [(10) Okay! We 'stayed' there…okay…not 'live'…stayed there for three days. [(11)He's…he has a problem with these…these s's (plural form).] [(12) Went there by the big car of my father…we went there in…, okay!] [(13)What is this? What does this tell me? This is going to be a U (band); can't avoid that. A U1? (long pause) U1? Maybe a U2?] Something like U1 lah!] [(14)… because…problem is the ideas…not coming across. I have to infer; 'prepared' as this is 'prepaid'. What does he mean by 'prepaid'? I have to read on to decide.] [(15) 'Live' there …I was thinking 'live' means 'stay' but here she has misunderstood the words 'live' and 'stay'… Okay!]

ER4/VP/Transcript/13.6.06 (Essay 1)

1.1 Data-examination: identify → (spelling error)
1.2 Data-examination: identify → (spelling error)
1.3 Data-examination: identify → (spelling error)
1.4 Data-explanation: read → [We live there for four day(s) three night(s)]
1.5 Data-examination: read → (I went there by the big car of my father)
1.6 Data-explanation: infer → (sentence structure/meaning)
1.7 Data –discrepancy-processing: recognition→ (wrong choice of word)
1.8 Data-hypothesis generation: trigger → (alternative word)
1.9 Data-evaluation: disconfirmation → (sentence meaning)
1.10 Data – discrepancy-processing: recognition→ (wrong choice of word)
1.11 Data – discrepancy-processing: recognition→ (plural forms)
1.12 Data – examination: identify → (preposition error/sentence structure)
1.13 Data-hypothesis generation: trigger→ […a U (band)]
1.14 Data-hypothesis generation: further specification → (clarity of meaning)
1.15 Data-hypothesis generation: further specification → (choice of word)

Table 4.5 ER4/VP/Transcript fragment and coding

On the other hand, the novice raters somehow seemed to have the tendency to focus more on data examination which to them was a more practical thing to do when assessing students' essays, followed by what they thought as sound data explanation of the surface-level errors. This is clearly illustrated in the transcript fragment and coding of NR 1 in Sample 2 (Table 4.6) and Sample 3 (Table 4.7).

Sample 2

This essay is about a family trip to the Damai Resort. The first sentence has a preposition error and a spelling error. There is another spelling error ('their' instead of 'there')

[(1) Basically, this student is writing about a story; something maybe (about) bringing a family (family trip) to the Damai Resort.] [(2) But even the first sentence itself has an error (preposition error). Like 'At' 27 December 2003, supposed to be 'On' December 27th, 2003, my family and I…(pause)] [(3) A lot of spelling error here; 'prepared' being spelled as 'prepaid'.] [(4)…em…hence wise (otherwise), the student is able to use past tense correctly… (pause). [Another (5) spelling error; 'there' becomes 'their'...]

NR1/VP/Transcript/15.6.06 (Essay 1)

1.1 Data-examination: read → (family trip to the Damai Resort)
1.2 Data-examination: identify → (an error (preposition error)
1.3 Data-examination: identify → (spelling error)
1.4 Data-meta-reasoning: experiential-memory → [(otherwise) the student is able to use past tense correctly]
1.5 Data-examination: identify → (spelling error)

Table 4.6 NR1/VP/Transcript fragment and coding (a)

Sample 3

There are hanging sentences, incorrect use of article (pronounce 'they'), incomplete sentence, wrong tense, direct translation from mother tongue, and spelling errors.

[(1) Sentence hanging(Dangling sentence)]… [(2) incorrect use of article (pronoun)…('They' instead of 'there')…] [(3) Some sentences are confusing with no subject (incomplete sentence)… (long pause)] [(4) Again tenses, supposed to be past tense. (long pause)] [(5) Confusion of words (wrong word choice) 'taken' a fish instead of 'caught' a fish.. (pause).] [(6) Confusing (distorted) sentence... (*We invited there chef to cooked it to become our food.)* [(7) Tenses ('feel' instead of 'felt') again]…. [(8)Confusing sentences]…. [(9)should be translated from one of the person's mother tongue…] [Yeah, it's (10) direct(ly) translated from the person's mother tongue, I think…] [(11) I suppose there are a lot of tense mistakes]… [(12) Spelling error as well…]

NR1/VP/Transcript/15.6.06 (Essay 1)

3.1 Data-examination: identify → (hanging sentences)
3.2 Data-examination: identify → [incorrect use of article (pronoun)]
3.3 Data-explanation: infer→ [no subject (incomplete sentence)]
3.4 Data-examination: identify → (tense/past tense)
3.5 Data-explanation: infer → [confusion of words (wrong word choice)]
3.6 Data-explanation: infer → [confusing (distorted) sentence]
3.7 Data-examination: identify → (tense)
3.8 Data-examination: identify → (confusing sentence)
3.8 Hypothesis-generation: trigger → [confusing (distorted) sentence]
3.9 Hypothesis-generation: generalization → (…should be translated from one of the person's mother tongue)
3.10 Hypothesis-evaluation: confirmation → [direct(ly) translated from the person's mother tongue]
3.11 Data-examination: identify → (tense)

Sample 3 (continued)
3.12 Data-examination: identify → (spelling error)

Table 4.7 NR1/VP/Transcript fragment and coding (b)

4.2.4 The Mental Model: Interpretation of the Expert and Novice Raters' Line of Reasoning

In order to be able to create a mental model from the line of reasoning, it is important for one to note that it is a functional abstraction of the assessing writing task given to raters which provides a deductive framework for assessing writing. This mental model contains and integrates conceptual knowledge, procedural knowledge, decision-making skills for reasoning, and strategies for assessing and giving feedback on students' writing.

In this study, a mental model was constructed based on the expert raters' line of reasoning. The line of reasoning (LoR) is '…a reasoning artifact that representing a complex set of related pathways of thought bound together' (Hassebrock & Prietula, 1992, p. 630). The knowledge states content of a LoR serves to discriminate expertise as well as individual differences between the expert and novice raters in assessing writing.

Writing is generally considered a complex, individualistic and idiosyncratic process. Thus, before one can assess writing, it is important for us to understand what is involved when a learner is engaged in the process of writing. When a learner writes, he/she undergoes an active thinking process, ass he/she comes to grips with the target language. As he/she struggles to put his/her thoughts down on paper, he/she would have to constantly search for the right words, the right sentence structure. In the process of doing so, he/she would discover something new to write about and a creative way of expressing these ideas. Thus, the writing process involves selecting

and organising experiences according to a certain purpose which requires active thought on the part of the learner. The learner would have to bear in mind his/her purpose, and select relevant information for that purpose and then would consider organising this information into paragraphs.

Thus, in assessing a piece of writing, a rater needs to be able to identify whether or not the students are in control of content, syntax, grammar, mechanics, organisation, word choice and choice of expression. In terms of content, students need to have the ability to think creatively and develop thought while excluding all irrelevant information. Thus, content would include relevance and originality of ideas, and also clarity of thoughts. In terms of syntax, a rater needs to be able to assess if students have the ability to compose a series of well-connected sentences that are grammatically and logically correct (Halliday & Hassan, 1976).

In the study by Vaughan (1991), he found that raters focused on a number of key linguistic features or elements when assessing second language writing ability: content, organisation and grammar, followed by handwriting and punctuation. The results of Vaughan's study imply that raters are not adhering to a single, internalised method for judging second language writing ability. In addition, it appears that despite the use of assessment guidelines, raters tend to rely on their own rating strategies when assessing writing samples which are deemed to be 'borderline' cases. A rater also needs to be well aware of students' stylistic skills, which is the ability to manipulate sentences and paragraphs, and use the language effectively. At the same time, a rater needs to be able to assess students' ability to write in an appropriate manner for a particular purpose with the ability to select, organise and order relevant information that clearly exhibits cohesion and unity within that piece of writing.

Based on the findings discussed in sub-section 4.2.3 (related to the lines of reasoning) two mental models were constructed and adapted from the general canonical model of assessing and giving feedback on students' writing given in Figure 3.3. This canonical model was earlier constructed to reflect the process of assessing and giving feedback on students' writing. It has provided a hypothesized qualitative description of the knowledge states and the conceptual operation that a rater may use in the analysis of a problem (analyzing error(s) in a student's writing).

The ER mental model constructed from the study is shown in Figure 4.1, while the NR mental model is shown in Figure 4.2. The difference between the two mental models is the emphasis given to the knowledge states and conceptual operations used during the assessing process. The differences are based on the analysis of aggregated data gathered from both categories of raters as presented as in sub-section 4.2.3(c).

The ER mental model shows that the expert raters focused certain key linguistic features or elements when assessing second language writing ability: content, grammar, vocabulary, and organisation, followed by mechanical aspects. In terms of the conceptual operators being deployed during the verbal protocol analysis, the priority list of the expert raters includes meta-reasoning, data examining, data explaining, data exploration, discrepancy processing, and hypothesis generation, followed by hypothesis evaluation and summarization.

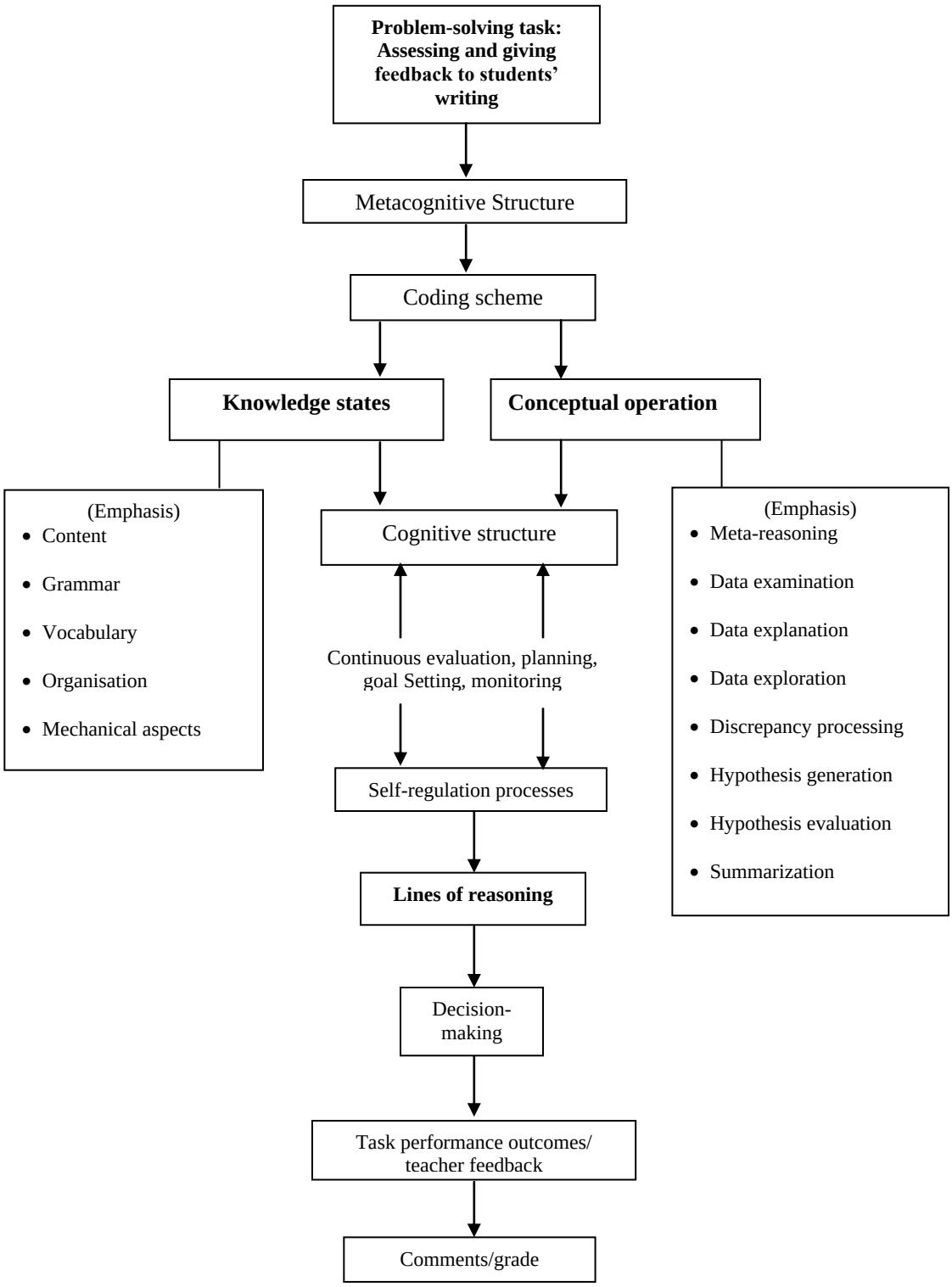

Figure 4.1 Expert rater mental model for assessing and giving feedback on students' writing

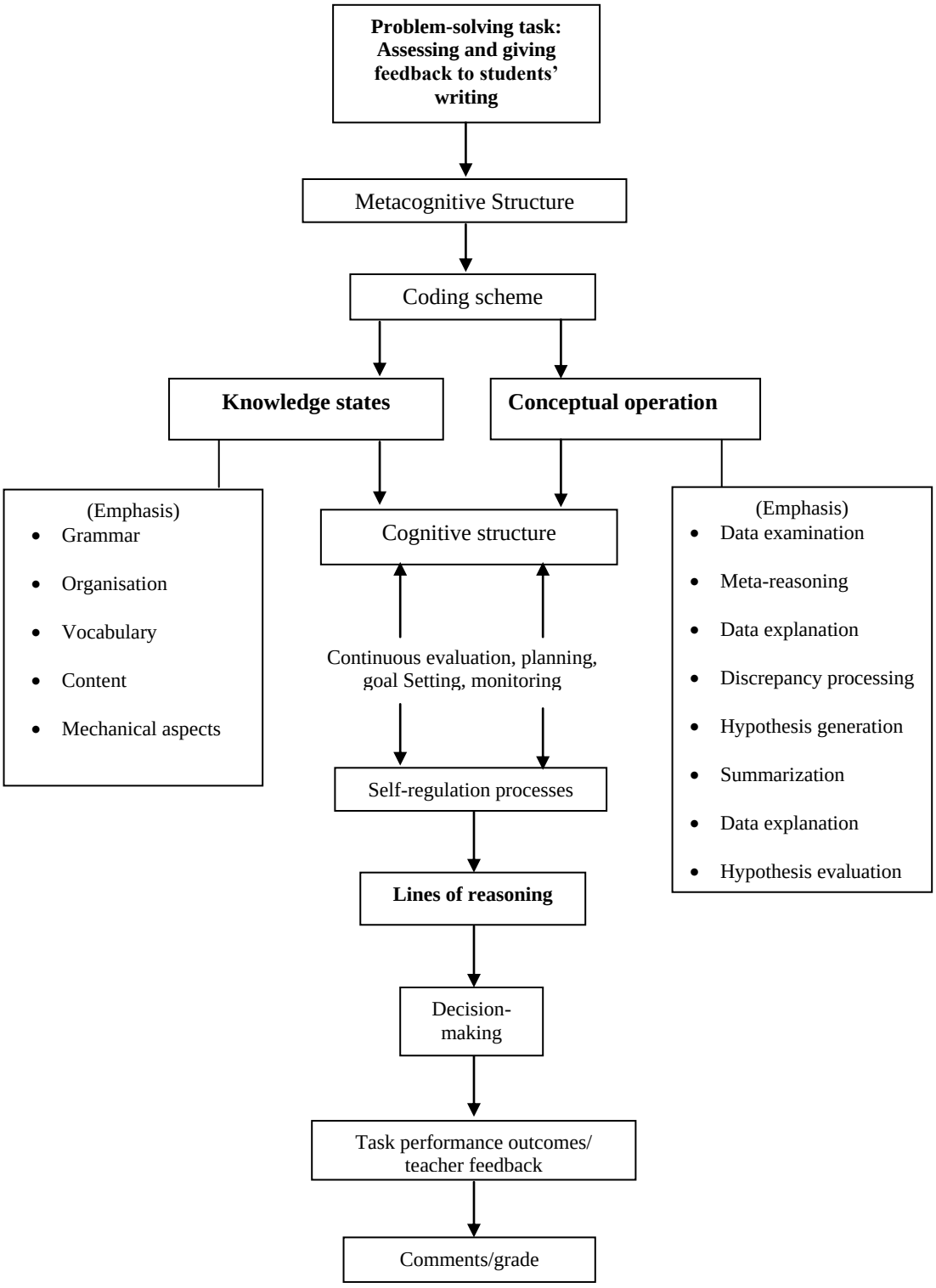

Figure 4.2 Novice rater mental model for assessing and giving feedback on students' writing

On the other hand, the NR mental model shows that the novice raters when assessing second language writing ability prioritise knowledge states like grammar, organisation, vocabulary, and content, followed by mechanical aspects. In terms of the conceptual operators being deployed during the verbal protocol analysis, the priority list of the novice raters includes data examination, meta-reasoning, data explanation, discrepancy processing, and hypothesis generation, followed by summarization and data explanation.

4.2.5 Using the Expert Rater Mental Model in Training Workshop: Novice Raters' Perception

This section will discuss how the course participants (Cps) or novice raters perceive the training workshop that was designed based on the ER mental model. At the beginning of the workshop, during the ice-breaking session, a majority of the novice raters thought that it would not be any different from any other courses that they had attended, whether it was at the school level or district level. They expected to be given inputs, including handouts that would serve as a guide if they need to apply in their classrooms whatever skill and knowledge that they were suppose to learn during the workshop.

During the workshop, the researcher designed sessions that basically required participants to get involved in a hands-on experience of assessing and moderating marking of sample writings (refer to Appendix 15). Discussions among the participants in smaller groups of five were focused on a process which was a new experience for both the participants and the trainer (researcher). There were evidently some challenges but also a developing awareness of what was needed to ensure that outcomes were productive for all participants.

(a) Novice Raters' Perception of the Knowledge and Skills Acquired through the Workshops

During the workshop, the researcher encouraged the novice raters to discuss issues concerning the pertinent knowledge states used by the expert raters, such as development of paragraph (ideas), variety of sentence structures, and wordiness. The novice raters had also explored the marking symbols and criteria needed in assessing writing, and had produced their own guidelines in giving feedback as part of an exercise to look back into their own prior knowledge and practice. During the workshop, it is worth noting that the participants were enthusiastic of learning the assessing strategy that the expert raters may use in assessing writing though they expressed their concern of lack of experience or confidence in assessing writing. This is very obvious through the feedback giving by the participants during the first session of the workshop when they were asked to list down what they brought along with them to the workshop. John, one of the participants, mentioned this when asked by the researcher how he felt at the beginning of the workshop as in Example 1:

Example 1

> I am willing to learn…I know I do not have much experience (in assessing writing) that I can share with others…I am certainly grateful with the knowledge and skill gained through the discussion.
>
> Transcript/Interview/CP2/12.6.08

The workshop was seen to provide support to teachers and helped with insights into the process of assessing and giving feedback on students' writing. This is evident in the teachers' comments. They reported that they had greater confidence and knowledge about assessing writing and this seemed to influence their classroom practice. For example, teachers commented on how they were using knowledge of key language terms and features, introduced by the trainer and clarified during the moderation process, when assessing their students' writing. This, they noted, had assisted them to make writing instruction more explicit for their students through

increased awareness of what they and students need to know about improving a piece of writing. John and Mary noted this in their response to the researcher's question on how the workshop helped them in their classroom practice, as given in Example 2 and Example 3:

Example 2

> I am more aware of the (language) terms and features related to good writing especially pertaining to organisation, content, grammar, mechanics (of writing) and vocabulary…it is so much easier to give feedback if you know what to look for and want to focus on. Most importantly…I am now more aware of the positive input that any constructive feedback can bring to improve my students' writing.
>
> Transcript/Interview/CP2/12.6.08

Example 3

> …before this, giving feedback, especially constructive ones seemed like a daunting task. However, this workshop taught me to be selective on the focus of my feedback…say for example, I may just want to focus on content like development of ideas or paragraph…err maybe clarity of ideas. Now I approach it (giving feedback) with the good intention of making improvements and creating better writing rather than criticising and judging…
>
> Transcript/Interview/CP3/6.7.08

Kim, on the other hand, focused less on the kind of feedback she could give to the students' writing and more on her increased awareness of the "style" of assessing expected by an expert rater. In particular, she provided the following response to the researcher's question about whether she thought her writing assessing skill had improved after attending the workshop:

Example 4

> OK, I think pretty well. OK, maybe not so much on how to give feedback, because that would depend on what students have written. But I have to say that this workshop didn't teach me that (giving feedback), it pushed me to about how to assess students' writing though. I expected the workshop to equip me on how to give specific response to the students' writing.
>
> Transcript/Interview/CP1/10.6.08

Here, the novice rater seemed to be suggesting that even if her giving feedback skill did not improve much, the workshop had at least increased her awareness of the how to assess students' writing.

Based on what is presented here as evidence, the researcher concludes that the workshop has succeeded in imparting some of the knowledge and skills used or emphasised on by the expert raters to the novice raters.

(b) Usefulness of Workshop

Another main theme that arose in the interviews concerned the novice raters perceptions of the usefulness of the workshop. All the participants who were interviewed gave a very positive response on the usefulness of the workshop in their own practice. Kim illustrated this in her response to an interview question as in Example 5.

Example 5

> I think I learnt a lot from the moderation process of marking and assessing the sample writing. It is easier with a small group because you can listen to each other giving comments on students' writing…small group means you got to know each other well enough to feel comfortable.
>
> Transcript/Interview/CP1/10.6.08

In the above example, characteristics of the group seemed to have affected the effectiveness of the moderation. As another participant, Mary, put it: '…big groups often meant no discussion…I mean there would be less opportunities for one to speak out his/her mind' (Transcript/Interview/CP3/6.7.08). In contrary to this, I noticed that Mary did not speak much during the group discussion. When I asked her about this, she replied, '…with teacher from other schools I was a bit shy with my opinion, but I learnt a lot' (Transcript/Interview/CP3/6.7.08).

As may be remembered, the novice raters only attended a workshop where they were given limited input due to time constraint. Many of the novice raters who were interviewed indicated that they wished they could have had more opportunities to attend similar workshops as they needed intervals to practice what they had learnt during the workshop. For example, in the exchange below, Chris provided the following information:

Example 6

> …and the other thing I would have liked is to get more opportunities to attend similar workshops…you know, I learnt something from this workshop but I need time to practise it. If there is a follow up, then it would be a good venue to share my experience with other Cps…
>
> Transcript/Interview/CP4/4.6.08

This request for additional workshop was made by other novice raters as well, as can be seen from the interview with Gerald in Example 7.

Example 7

> Well, uhm, I think it's a good idea to have more workshops like this. I'm not sure if having just one workshop would be enough to equip me with the necessary skill to assess students' writing effectively. It would be very beneficial for people like me to attend a series of similar workshop. Otherwise, it would be a good idea to stretch the two-day workshop to a five-day…
>
> Transcript/Interview/CP/3.6.08

Towards the usefulness of the knowledge and skills they acquired through the workshop, the novice raters generally held favourable opinions, indicating that the knowledge and skills helped them improve the quality of their assessing and giving feedback on students' writing, as was suggested by Mary's comments on her perception of the workshop in Example 8:

Example 8

> Uhh, it was good, I learned in that way. Also, for me it was very important to get familiar with effective way to assess students' writing. This workshop has given me the opportunity to explore how assessing writing can be done effectively and consistently. It also helps me in giving better comments.
>
> Transcript/Interview/CP3/6.7.08

However, it could be suggested at this point that the novice raters were merely giving their 'ideal' response. In other words, there could have been a halo effect, with the participants providing the information that they believed the researcher was expecting (Mackey & Grass, 2005). In addition, it could be suggested that these novice raters, who voluntarily gave their time to speak to the researcher, may have had more favourable views about the workshop than those novice raters who did not volunteer (i.e. a Hawthorn effect, Mackey & Grass, 2005). However, the novice raters were not completely uncritical of the knowledge and skill they received. In Example 9, for example, John indicated that he disagreed with the kind of learning she gets from the workshop:

Example 9

> …I mean some things were really helpful, you know I was completely in agreement with. You know, like the way how we moderate the marking of scripts…I seldom do this in my school. This is something different. But then, how long can I sustain this kind of skill when I get back to my school.
>
> Transcript/Interview/CP2/12.6.08

Comments such as those in Examples 5 and 9 only provided a partial glimpse of the "hidden transcript." In light of the fact that the researcher was also the trainer of the workshop, the novice raters may have felt reluctant to express their concerns more directly, and most novice raters apparently opted to avoid giving less positive comment. Nevertheless, the fact that this tension was raised during the interviews suggested that such workshop needed to be carefully designed and monitored on an on-going basis to ensure that the novice raters' needs were being met. This revelation

posed a considerable challenge in workshop design and implementation: the trainer who implemented it needed a considerable effort to gain a background in the novice raters' prior knowledge, working in close collaboration with experts in the subject matter (expert raters), and crucially listening to the feedback of the novice raters - such as those reported here.

(c) Novice Raters' Confidence in Applying the Knowledge and Skills Acquired through the Workshops

One of the major themes that emerged from a reiterative and inductive analysis of the interview transcripts concerned participant's confidence in applying the knowledge and skills acquired through the workshop. As can be seen from the comments below, the novice raters did not conceptualise their improvement in terms of a greater mastery of error identification, but rather focused on the gains they had made in assessing and giving feedback on writing. In Example 10, Mary articulated her view on what she gained from the workshop.

Example 10

> I think that the assessing skill now has improved. I no longer rely on just error identification to decide on the kind of marks I will give to the students' writing. That is very tedious. In fact, this workshop has helped me to look beyond just error identification…more to content analysis, development of paragraphs and completeness of the piece of writing. I think I will continue to sharpen my skill through my own classroom practice.
>
> Transcript/Interview/CP3/6.7.08

Other novice raters phrased this sentiment in terms of their assessing skill more closely resembling the mental model used by the expert raters. For instance, in Example 11 from Chris and Example 12 from Gerald, they stated:

Example 11

> I managed to get the overall gist of the essays by looking at how paragraphs are developed, and to draw some nice comparisons among them and to come up with I think some strong conclusions, but it was grounded on the facts of very good writings I had found (among the essays that were being assessed), and I am pleased with that. Uh, to be able to do something that when I read it (students' writing) I thought "Oh! This really looks like the decisions that may have been made of seasoned (examination) markers.
>
> Transcript/Interview/CP4/4.6.08

Example 12

> I really liked how I made my decision, I mean, in my lack of experience in assessing writing it seems like a true assessing skill and something that is good…I would say that, after completing a task (assessing writing), I have that kind of feeling that I have made a thorough analysis of its completeness to decide what kind of weaknesses and strengths it has against real good writings.
>
> Transcript/Interview/CP3/3.6.08

Kim, on the other hand, was rather sceptical on how well she can really apply what she learnt during the workshop in her practice. In her response to the interview question about whether she thought her learning during the workshop would help her in her practice in her school, she articulated in Example 13 that:

Example 13

> I learnt something here (during this workshop) but I think I need more practice and participation in a workshop like this to be really good in assessing writing…however, I am willing to learnt till I really acquire the skill.
>
> Transcript/Interview/CP1/10.6.08

4.3. Discussion on Research Findings

The findings that have been presented here provide valuable input in answering the research questions outlaid in this study. While it is lucid that differences between the novice and expert raters in assessing writing do exist, discussion on these results based on information from other literature is needed to further understand the phenomenon under studied. The following sections will answer each research question accordingly.

4.3.1 Research Question 1: What are the knowledge states used by the expert and novice raters in assessing and giving feedback on students' writing?

This sub-section discusses the knowledge states used by both the expert and novice raters in assessing and giving feedback on students' writing.

Based on the discussion in sub-section 4.2.1, it is obvious from the protocols that, as in the study of expertise in other domains, the content and organization of knowledge states played the major roles in differentiating the expert raters from the novice raters (Hassebrock & Prietula, 1992). As clearly shown in Table 4.1, the expert raters were rapidly able to construct a rich mental framework to relate their knowledge to specific writing features and to anticipate what was to come through the conceptual operation they had employed to effectively assess the writing in a short period of time. This could be largely due to the experience that they had and a collection of knowledge that they can draw upon (Barkaoui, 2007; Condon, 2009). As such, there was rapid pattern recognition of knowledge states to complement the conceptual operators used during the problem-solving task (Hassebrock & Prietula, 1992).

As in most domains, such as mathematics (Suto & Greatorex, 2008) and physics (Simon & Simon, 1978; Larkin, 1981), the expert raters' greater knowledge resulted in the problem being recognized as a familiar kind, and there was rapid task solution. The novices, on the other hand, had demonstrated quite a limited recognition of knowledge states and had spent more time on each writing as compared to the experts. They did not have the organization of schemata to allow effective processing, which impeded their speed in assessing each task. The expert raters, as the researcher has indicated, had great domain-specific information, and this information was highly organized and conceptually integrated (Hassebrock & Prietula, 1992). These findings are shared by Sakyi (2000) who found similar patterns in his comparison between the

expert and novice raters. Organized knowledge appeared to account for the rapid pattern recognition and categorization apparent in their performance. Moreover, experts raters used their knowledge in qualitative reasoning as well as quantitative reasoning - with notable facility.

4.3.2 Research Question 2: What are the conceptual operators used by the expert and novice raters in assessing and giving feedback on students' writing?

This sub-section discusses the conceptual operators used by both the expert and novice raters in assessing and giving feedback on students' writing. Based on the discussion in sub-section 4.2.2, expert raters had a rich stock of schemata to frame the problem (Hassebrock & Prietula, 1992) as shown in Table 4.3. It is obvious that the expert raters had used more varied forms of conceptual operators as they mined their knowledge resources to provide a deeper, richer exploration of the stimulus state of the task, thus they were able to give quality feedback to students' writing. Their emphasis on meta-reasoning facilitated the process of assessing a piece of writing in a more efficient way as students' errors could be easily identified and hypotheses generated during the rating or marking process could be evaluated and summarised more effectively.

The novice raters, on the other hand, had deliberately constructed the representation of the task in a slow, step-by-step manner but not explore the writing in any depth after they had some grasp of the errors identified in it. This is somehow similar to the findings of a study on expertise in interpreting congenital heart diseases in the medical domain (Hassebrock & Prietula, 1992), in which the novices were consumed by the task of building up a representation of the subject matter, whereas the experts could focus on higher level of interpretation.

4.3.3 Research Question 3: How does the expert raters' lines of reasoning differ from the one used by novice raters to represent the knowledge states and conceptual operators in assessing and giving feedback to students' writing?

This sub-section discusses how the exporter raters' lines of reasoning differ from the one used by the novice raters, especially in terms of knowledge states and conceptual operators represented in assessing and giving feedback on students' writing.

Based on the discussion in sub-section 4.2.3, the analysis indicates that expert raters tended to conceptualize assessing activity as a macro-strategy and as a task that was predominantly undertaken as a post-textual production reviewing activity. This is similar to the findings by Eckes (2008), who discovers that expert raters tend to have enough knowledge to articulate their reviewing process of a written work.

The analysis of the findings also indicates that expert raters engaged in multiple reviewing activities during assessment of writing, including many revisions that were not concerned with simple matters of surface accuracy, and many expert raters were able to talk about these perceptively and with insight. Cumming, Kantor and Powers (2002) further support this by revealing in their study that ESL and EFL expert raters attended more extensively to language, rhetoric and ideas as a whole rather than focusing on a specific element as in the case of novice raters.

With regard to the expert raters, Chi et al. (1988) states that expertise is characterized by a fluid, automatic process, frequently inaccessible to conscious reflection. It is only when comprehension breaks down that consciousness is triggered (Crisp, 2008; Afflerbach & Johnston, 1984). In this study, after examining the expert raters' protocol from the pilot study and as well as the final study, the researcher decides that coding the expert raters' 'think-aloud' is indeed productive in understanding how they

differ from the novice raters in terms of cognitive processes as they assess students' writing.

This study illustrated that there are differences in how the expert raters and novice raters assess writing. Clear trends emerged in terms of knowledge states and conceptual operators being deployed during the verbal protocol analysis task of assessing sample writings. Similar to the solution of elementary physics problems (Simon & Simon, 1978; Larkin, 1981), differences in qualitative as well as quantitative reasoning distinguish the experts' from novices' performances. The two disparities between the expert and novice raters are easily noted. The most obvious disparity is time-to-solutions, as noted by the researcher during the video recording of the VPA, the speed with which a problem was solved. The skill of the individual determined how rapidly a solution was achieved. Expert raters were generally more efficient in searching and representing their solution space and, as a result, may perform twice or thrice times faster than novice raters when tackling standard writing assessing problems (Cumming, Kantor & Powers, 2002).

4.3.4 Research Question 4: How can the expert raters' line of reasoning be interpreted in the form of a mental model that can be used to help novice raters in assessing and giving feedback on students' writing?

This sub-section focuses on how the expert rater's line of reasoning can be interpreted in the form of a mental model. In this study, this mental model is a problem representation of assessing writing task that enables a person to actively gather information, make inferences, anticipate outcomes and make plans for future decision-making (Newell and Simon, 1972), apart from providing a source of information (Qin and Simon, 1995) that can be manipulated to make predictions and inferences. In this study these processes were imbedded in the problem-solving task

as the expert raters' make use of their prior knowledge states and at the same time engaging the various conceptual operators to assess sample writing effectively.

In this study, the findings which are based on aggregated data shows that the expert raters focused on a number of key linguistic features or elements when assessing second language writing ability: content, grammar, vocabulary, and organisation, followed by mechanical aspects. In terms of the conceptual operators being deployed during the verbal protocol analysis, the priorit y list of the expert raters includes meta-reasoning, data examining, data explaining, data exploration, discrepancy processing, and hypothesis generation, followed by hypothesis evaluation and summarization. However, it needs to be noted that this does not mean that the expert raters adhered to a single, internalised method for assessing second language writing ability. As individuals, expert raters seem to rely on their own rating strategies when assessing writing samples.

Based on the discussion of the findings in sub-section 4.2.5, comments given by both the expert and novice raters during the VPA only provide a partial glimpse of the hidden 'transcript.' In light of the fact that the researcher was present during the course of the VPA, despite his non-involvement, raters may have felt reluctant to express their concerns more directly. A close scrutiny of the situation leads the researcher to believe that the novice raters, for instance, apparently had tried to present their VPA in a manner that would not reflect their lack of competence in assessing writing.

On the other hand, the expert raters may also camouflage their feeling of short-changing the researcher's expectation while doing the task. Nevertheless, the fact that

this tension was raised – even directly – in both the VPA and interviews suggest that adjunct workshops or courses such as 'Assessing Writing for Beginning Teachers Workshop' need to be carefully designed and monitored on an on-going basis to ensure that the students' needs are being met. Based on the qualitative analyses in this study that provide the triangulation and support, a more holistic picture could be portrayed by exploring and understanding the association between the novice and expert raters' lines of reasoning that helped in the construction of a comprehension mental model that could be used to help novice raters acquire the necessary skill to assess students writing effectively.

4.3.5 Research Question 5: How can the novice raters approximate the expert raters in assessment behaviour through training based on the mental model of the expert raters?

This sub-section discusses the novice raters' perceptions of the training workshop and how it affected their practice in the school environment, pertaining to their learnt skills of assessing and giving feedback on students' writing. Based on the discussion of the findings in sub-section 4.2.5, the analysis suggested that at the outset of the workshop, novice raters were fairly satisfied with the workshop and felt reasonably confident about assessing writing. One of the obvious implications of the intervention workshop is an awareness that teachers need to collaboratively reflect on their professional needs and current understanding on assessing writing. The workshop on assessing writing had given a lot of focus on the discussion of assessing and giving feedback on students' writing, giving opportunity to teachers to investigate their own practice and deepen their understanding of the assessing process.

During the workshop, decision about the assessing process was negotiated by the novice raters working on the same task. In addition, there were many informal

discussions as well. This process apparently had contributed to novice raters to become skilled enquirers who could improve their own practice through collaborative mode. Thus, the workshop had provided novice raters with opportunities to learn the skilled required in assessing their students' writing more effectively.

The intervention workshop has also drawn a realisation upon us that in order to assess students' writing effectively, teachers need to have an explicit knowledge of the knowledge states (Hassebrock & Prietula, 1992) such as grammar, mechanics, content, organisation and vocabulary, and the conceptual operators that will lead to better decision making that benefits students. Furthermore, teachers need a meta-language (language use to analyse language) in order to describe and discuss language and able to include it as a natural part of assessing writing. This involves the deliberate control of what to think about and how to think in order to maximise progress and minimize error in any problem-solving task one is engaging in (Kuhar, 1998).

The ability to describe language in terms of text and grammatical features is invaluable because it enables teachers to focus precisely on the meaning by which writers (the students) shape and manipulate their assessors' thoughts and feelings. In this study, after a brief exposure to some of the assessing strategies in a simulated training/workshop, the novice raters subsequently used more conceptual operators that inferred relationships among information from the students' written work.

4.4 Chapter Summary

In this chapter, the quantitative results and qualitative analyses of the study based on the research objectives and research questions as stated in Chapter One were

presented. The data presented here shed light on many important factors in both the types and modes of assessing strategies used by raters. The knowledge states and conceptual operators employed by the novice and expert raters were identified and discussed in sub-section 4.2.1 and sub-section 4.2.2. In addition, this chapter also presented the results of the investigation regarding the expert and novice raters' lines of reasoning in sub-section 4.2.3 and how they differed among these two groups of raters.

In relation to the mental model that was being constructed from these lines of reasoning, Sub-section 4.2.4 presented the qualitative analyses that provided the triangulation and support to the construction of this mental model. Therefore, a more holistic picture could be portrayed by exploring and understanding the association between the novice and expert raters' lines of reasoning. Later, in sub-section 4.2.5, the findings from the trialling out the conceptual model of the expert raters in the form of training sessions (workshops) with a group of beginning teachers (with some experienced teachers in attendance) was also presented to gain insights on the participants' perception on the effectiveness and implications of the intervention training. Section 4.3 gave a discussion of the findings of this study according to the four research questions. The findings were discussed in lights of previous literatures and related findings of past studies. Section 4.8 ends this chapter with a summary on what had been presented and discussed in this chapter. The conclusion, implications and recommendation of the research will be discussed in the next chapter.