

## **CHAPTER 2**

### **REVIEWS OF RELATED LITERATURE**

#### **2.1 Chapter Overview**

The literature review for this study as presented in this chapter serves three purposes: (a) to construct a theoretical framework for the study, (b) to review related literature pertaining to the two main variables of this study – the knowledge states and conceptual operators, as they are extensively used in the data analysis, and (c) to review related literature in the methodology to be used in this study.

#### **2.2 Constructing a Theoretical Framework for the Study**

The theoretical framework of this study was constructed based on the discussion on the writing assessment and evaluation in second language teaching. The discussion on writing assessment was focused on the four types of product measures; the holistic scoring, primary trait scoring, analytic scoring, revision analyses and error analysis. On the other hand, the discussion on writing evaluation was focused on the three general types of rubrics used in evaluation; the non-weighted rubric, the weighted rubric and the holistic rubric. Knowledge of these writing assessment and evaluation methods may have distinct association with how a teacher/rater provides feedback on students' writing to improve their writing performance, especially the much needed knowledge on the factors that contribute to good writing.

### **2.2.1 Writing Assessment in Second Language Teaching**

Assessment is generally accepted as a method to measure students' progress and attainment. Specifically, Smith (1999) defines assessment as the set of processes through which we make inference about learner's learning process, skills, knowledge and achievements. Popham (1995), on the other hand, mentions that assessment is a formal attempt to determine students' status with respect to educational variables of interest. It is a deliberate effort to determine a student's knowledge, skills or attitudes.

He also maintained that the traditional reasons teachers assess students are:

- i. To diagnose students' strengths and weaknesses
- ii. To monitor students' progress
- iii. To assign grades to students
- iv. To determine instructional effectiveness

For Tindal and Marston (1990) assessment is

...testing, appraisal, diagnosis, measurement, performance review, and evaluation are some of the labels used in education...the process that helps determine children's skills and knowledge...is the systematic process we use to gather data that allow us to instruct students more effectively. (p. 27)

In education, assessment is seen as compulsory as other components such as policy, planning, curriculum development, management, teacher training etc. This is proven when numerous studies have been done overseas or locally on how significant assessment is in education. Campbell and Evans (2000) for instance, have stressed on the importance of consistency in the assessment of pupil achievement with the identified learning outcomes/ instructional goals at the intended level of performance. They stated, "As part of instructional planning, teachers must decide whether identified instructional goals lend themselves to assessment through an affective, a cognitive or a psychomotor domain" (Campbell & Evans, 2000, p.11).

Rabiah (1994) in her study has proposed to disclose the current situation of English language teaching and assessment across Sarawak. She has discovered that the curriculum design needs to be changed. Teachers should be made to understand that the syllabus and national public examination have influence on classroom teaching testing processes. Pophams (1995) further postulates that educational assessment results influence public's perceptions of educational effectiveness, help evaluate teachers and clarify teachers' instructional intentions. Ponnudurai (1996) claims that the basic purpose of examination is to reflect the educational aims of the system and it is agreed that high quality exams could read to high quality learning. These studies proved that assessment is highly regarded as a tool to ensure effective teaching and learning in schools.

Sommer (1989) defines assessment as the process of finding out who the students are, what their abilities are, what they need to know, and how they perceive the learning will affect them. In second language writing, Sommer (1989) further distinguishes assessment from evaluation when he says that assessment takes place at the outset of the writing course, whereas *evaluation* describes on-going activities that eventually provide closure in the writing course. Assessment in writing is summative as it is used to discover the overall quality of writing, and is used for large scale evaluation to determine how much the students know. Hout and O'Neill (2008) elaborate extensively on the nature of writing assessment in ESL and EFL contexts. As such, assessment places the need of the students at the centre of the teachers' planning, which can be direct or indirect. Direct assessment reflects the classroom teaching of writing. This type of assessment gives the opportunity to students to show how they can organize, compose, argue, and use a variety of words and sentences. On the other hand, indirect assessment is objective and involves editing.

The four types of product measures that provide a broader assessment of writing are holistic scoring, primary trait scoring, analytic scoring, revision analyses and error analysis. In the following section, the discussion covers the advantages and disadvantages of these text-linguistic measures, which helps to outlay the foundation of the study's framework.

#### **a) Holistic Scoring**

In the 1960s, writing assessment was dominated by indirect tests of writing ability such as multiple-choice examinations of grammar, whereas the 1970s saw an increasing acceptance not only of direct testing (writing test) but also of a view of writing as a communicative act and as a recursive, nonlinear process which could not be taught or learned as a sequence of discrete stages.

Accordingly, holistic scoring – assigning a single score to an essay based on the reader's quick overall impression of the essay and how well it communicated the writer's ideas – became the most widespread assessment method and remained so until the mid-1980s (Cho, 2003; Eckes, 2008; Hamp-Lyons, 2002; Yancey, 1999).

According to Wolcott and Legg (1998), scoring holistically means that:

...an essay is not evaluated in terms of its specific features, for instance, its development, creativity, sentence structure, or mechanics; neither are ratings derived by mentally adding together scores for the individual features. Instead, a paper is scored in terms of the overall impression – the synergistic effect – that is created by the elements working together within the piece. (p. 71)

Holistic scoring may involve rank-ordering, that is, assigning the best papers to the highest scores and the worst the lowest (Cooper, 1977); however, scores are commonly anchored to a set of external criteria, such as a scoring rubric, which guide

trained raters in making their decisions (Arter & McTighe, 2001; Bailey, 1998). For example, in the holistic rubric for the continuous writing section provided by the Malaysian Board of Examination (2006) for the Malaysian Certificate of Education or *Sijil Pelajaran Malaysia (SPM)* (see Appendix 10), there is a six-point scale or range mark addressing various aspects of the quality of the essay, including language or grammar, sentence structure, vocabulary, punctuation, spelling, paragraphing or organization, and content. Rubrics, however, may range anywhere from zero to fifty points; an even number of points is commonly recommended so that readers must be forced to decide whether the essay belongs in the top-or the bottom-half of the group.

Holistic scoring gives an overall evaluation of writing. Teachers usually read student's for general or whole impression, and based on this impression they sort out essays into three, four, five, or six categories from very good to poor. Then the essays in each category can be awarded a numerical score or letter grade. Every aspect of the composition, both content and mechanical considerations, affects the teacher's response, but none of them are specifically identified or directly addressed using a checklist. Instead, the focus is on overall writing performance.

In this method, the students will not know their strengths and weaknesses as grading are given as a whole. Holistic scoring is mostly used in large scale assessment. Although this approach is fast and efficient to judge overall writing performance without emphasis on any particular writing skill, it is not an appropriate measure to use when teachers want to assess how well students have used a particular writing form or applied specific writing skills in a composition. The main disadvantage of this approach is that teachers may unknowingly place too much emphasis on mechanical

correctness, especially, spelling, grammar usage, and hand-writing and therefore bias in their assessment (Searle & Dillon, 1980).

Holistic scoring has been praised for rewarding students on what they do well (Wolcott & Legg, 1998). Holistic scoring is also the least time-consuming of all the assessment methods, as it involves assigning only a single score. For this reason, holistic scores have been commonly employed in L1 (and to the lesser extent, L2) writing studies as a means of measuring the impact of a particular form of pedagogic intervention, such as the use of various types of feedback.

However, holistic scoring has also been extensively criticised. For example, holistic scores have been found to correlate significantly with certain features of the text, including handwriting and text length (Steward & Grobe, 1979; Nold & Freedman, 1977), the number of grammatical errors (Sweedler-Brown, 1993; Homburg, 1984), Lexical features (Engber, 1995; Grobe, 1981), content and organization (Freedman, 1979) and rhetorical features (Tedick & Mathison, 1995), raising the question of whether holistic scorers actually assess the essay as a whole, or whether their decisions are guided by a relatively small set of criteria. Other researchers have criticized holistic scoring on the grounds that even when a scoring rubric is provided to focus the raters' attention on multiple aspects of the text, this in itself disconnects the holistic rater from the whole or gestalt of the text, creating an unnatural relationship between the text and the reader which distorts the final judgments provided to the text (Elbow, 1996; Huot, 1993, 1990).

Holistic scoring has also been criticized for obscuring the source of the score (Hyland, 2002; Weigle, 2002; Cumming, 1997), preventing students, teachers and researchers

from gaining a deeper understanding of the particular strengths and weaknesses of the essay. This is a particular concern when dealing with L2 learners, as these writers tend to exhibit more variation across the different subcomponents that make up the totality of writing skill (Hamp-Lyons, 1995). For example, an L2 writer may have solid control of grammar and mechanics, but at the same time, be aware of the rhetorical conventions that guide essay organization. Thus, when his/her essay receives a '3' on a scale of 1 to 5, we are left wondering whether '3' is an average of the high and low elements of the essay, whether '3' is indeed a true representation of the writer's abilities, and what exactly caused the writer to receive that score. As Hamp-Lyon (1995) notes "a holistic scoring system is a closed system, offering no windows through which teachers can look in and no access points through which researchers can enter" (p. 760).

Furthermore, even though researchers like Bailey (1998), and Wolcott and Legg (1998) have claimed that holistic scoring facilitates high inter-rater reliability rates (provided that raters have received training), other researchers have claimed that the field has experienced widespread methodological confusion with respect to inter-rater reliability rates (Polio, 2001). At the same time, though, other researchers have claimed that too much attention has been paid to inter-rater reliability and not enough to *validity* - that is, whether the holistic scale truly assesses the writing ability of the student (Huot, 1990). As Charney (1984) notes,

...the validity of holistic scoring remains an open question despite such widespread use[;] the question of whether holistic ratings produce accurate assessments of true writing ability has very often been begged; their validity is asserted, but has never been convincingly demonstrated. (p. 68)

Possibly due in part to these criticisms, it appears that no study investigating feedback and L2 writing has relied on holistic scoring as the sole measurement of writing

quality or change across drafts. Those who have used holistic measures have done so in conjunction with other assessment measures (Chandler, 2003; Blain, 2001; McGroarty & Zhu, 1997).

For example, Berg (1999) used holistic scoring in conjunction with revision analysis to measure the impact of training in peer revision on the drafts produced by forty-six university-level ESL students. She found that the group who had been trained in peer response received a significantly higher difference score (the holistic score of the second draft minus that of the first draft) than did the untrained group. Although the use of the holistic score did not allow her to determine which aspect(s) of the students' writing improved, the general improvements seen helped Berg to conclude that peer training did significantly benefit the L2 writers.

In another study to use a combination of measures, Prater and Bermúdez (1993) used fluency measures and holistic scoring to assess writing of elementary school ESL students. Writing in small group and receiving feedback from their peers on their drafts. The assessment evidenced greater pre test-post test gains in the quality of their compositions than did students who worked individually and only received feedback from the teacher. They found that even though the peer feedback group received significantly higher ratings on two out of the three fluency measures (number of words and number of idea units), there were no significant differences between groups in terms of the final holistic score received. Prater and Bermúdez (1993) attributed this to the short-term nature of the study, saying that "it is likely that one month is not a sufficient length of time to produce significant differences in overall quality of writing" (p.106) – a possibility that has been raised by other researchers as well (e.g. Chandler, 2003; Hillocks, 1986).

In addition to its reported lack of ability to measure short-term changes in writing quality, holistic scoring has also been criticized for its inability to detect differences between groups at close proficiency levels. Polio et al. (1998), for example, noted in their decision to rely on objective measures of linguistic accuracy rather on holistic scales that the later “were not fine-grained enough for use on a homogeneous population, that is, on students who have placed into the same class” (p.52). Due to the fact that these scales do not provide detailed diagnostic information on particular aspects of the students’ writing (such as grammar, organization, and so on), holistic scales are generally used for measuring differences between groups at a wider range of proficiencies.

Holistic scoring can help researchers delineate large differences between groups, and thus may be a useful time and cost-effective method for determining whether two groups are roughly at the same proficiency level or not. However, as this method cannot illuminate *how* the groups differ (for example, was one superior in grammar? another in organization?), it may have less utility when investigating finer-grained differences between groups at approximately the same level of proficiency.

#### **b) Primary Trait Scoring**

In primary trait scoring, the focus is on whether or not students have incorporated specific traits or qualities in their writing. Developed partly in response to the limitations of holistic scoring, primary trait scoring is a form of criterion-based assessment in which one trait of the writing (e.g. descriptive, persuasive arguments or organization) is chosen and then evaluated holistically (Wolcott & Legg, 1998).

The underlying assumption of this method is that different types of writing require different types of skills and strategies. Success in writing a descriptive essay, for example, will depend much more on the writer's ability to provide a detailed, example-rich description than on his or her use of persuasive arguments. Thus, in primary trait scoring, the evaluator begins by reflecting on the purpose of the writing, as well as on the types of rhetorical strategies, kinds of information and methods of presentation that would be important in carrying out the task successfully (Odell & Copper, 1980). The next step is to identify the most important feature – the primary trait – that would characterize a successful essay, given its purpose, audience and genre. The essays are then evaluated and given a single number, which represents the quality of the paper according to that single trait (White, 1986).

Primary trait scoring has been praised for “giv[ing] a sharper view of the complex of particular skills required to do a given task, and therefore increas[ing] the likelihood that we will be able to identify strengths and weaknesses precisely” (Lloyd-Jones, 1977, p. 47). Furthermore, given the fact that skill in writing one type of essay (e.g. description) does not necessarily transfer to skill in writing another (e.g., persuasion), primary trait scoring can help identify a student's strengths or weaknesses on a particular writing task and thus has the potential to provide more detailed diagnostic feedback (Wolcott & Legg, 1989).

However, this type of scoring has been labelled as “reductionist,” as it collapses the multifaceted nature of the writing into a single trait, thus hindering researchers and teachers from assessing the totality of skills that are involved in writing (Wolcott & Legg, 1998). Furthermore, this restricted focus may also pose problems for the evaluators themselves, as they may have difficulties in focusing on one trait of the

paper to the exclusion of others (Wolcott & Legg, 1998). Researchers have in fact suggested that such a focus is impossible. Hamp-Lyons and Henning (1991) note that facets are not independent but are interwoven throughout the text, so that readers' judgements of one facet are influenced by the quality of others.

In a study done by Ashwell (2000), the researcher employed multiple measures to investigate the timing of form form-focused and content-focused feedback, including a primary trait on the content of the essay. The researcher developed a primary trait rubric which provided descriptors for different levels of quality concerning the content of the essay. For example, the highest ranking category described the content of the students' essay as "display[ing] an ability to communicate with few or no difficulties for the reader," while the lowest described the writing as "display[ing] little or no ability communicate...[and] represent[ing] a totally inadequate answer" (Ashwell, 2000, p. 254). With this scale, he found that three out of the four groups of students received significantly *worse* content scores on their second drafts than on their first. Between the second and third drafts, however, all groups received slightly better, though not statistically, content scores. Although Ashwell suggested that the amount and type of feedback given may have caused the students to prioritize form over content, due to the fact that inter-rater reliability rates were quite low (less than 70 percent), it may be the case that the rubric was not an adequately valid measure of the quality of the content in the students' essays.

If carefully designed and piloted, primary trait rubrics may be a useful method for assessing discrete aspects of the learners' writing, such the quality of the content in their essays. Nevertheless, as writing is a multidimensional skill, involving a number of sub skills working in tandem (e.g., grammar, organization, argumentation skills,

discourse), such rubrics may also act as blinders, obscuring other aspects of change in the students' writing skills.

### **c) Analytic Scoring**

A more common assessment procedure in studies investigating feedback and writing quality is the analytic scale, also commonly known as multiple trait scoring. In this procedure, a set of traits is chosen (e.g., grammar, organization, content) and then evaluated individually (Brown & Bailey, 1984; Hamp-Lyons, 2001, 2003). These scores are then summed to arrive at the final score; alternatively, a weighted sum may be used to represent the relative importance of the various features in the final score (White, 1986).

In analytic scoring, teachers score writings against a range of writing skills. This form of assessment is most appropriate when teachers want to compare student's writing to a standard of excellence. Diederich (1974) developed an analytic scoring system for high school and college students that divided writing performance into two main categories, general merit and mechanics, and he identifies several specific traits related to each category. The specific traits for general merit are ideas, organization, wording, and style. The specific traits related to mechanics are usage and sentence structure, punctuation and capitalization, spelling, and handwriting and neatness. Percentage values could also be assigned to each category to determine a grade. Perhaps the most significant drawback of this system is that equivalent weight is given to the two categories even though writing educators recommend that greater emphasis be given to content.

Analytic scoring such as that used by the ESL Composition Profile has been the recommended scoring procedure for use with L2 writers, as it gives more-in-depth information about the writers' particular strengths and weaknesses (Bacha, 2001; Hamp-Lyons, 1995; Connor, 1991; Hamp-Lyons & Henning, 1991) and has been fruitfully employed in a number of studies investigating the relationship between feedback and L2 writing (e.g., Blain, 2001; Saito & Fuita, 2004; Paulus, 1999). Paulus' (1999) analytic rubric, for instance, (which included categories on organization/unity, development, cohesion/coherence, structure, vocabulary, and mechanics) allowed her to detect significant differences in the quality of writing produced by two groups of students, one receiving peer feedback and the other teacher feedback.

Analytic scoring, however (as with all other forms of writing assessment), has not escaped criticism. First of all, it is problematic to identify *which* sub skills should be included in the scoring rubric. Arguably, writing is composed of a wide variety of skills: spelling, punctuation, grammar, local grammar, local organization, global organization, persuasiveness, style, humour and so forth. Which ones are the most important to assess for a particular group of learners engaged in a particular task? Will different tasks and/or writers call for different sub skills? Or will the sub skills that are identified for one type of writing be irrelevant for another? The lack of agreement in the profession on what these sub skills are have suggested to some "that writing remains more than a sum of its parts and that the analytic theory that seeks to define and add up the sub skills is fundamentally flawed" (White, 1986, p. 123).

Nevertheless, as L2 writers often exhibit considerable variation across writing sub skills, the use of an analytical rubric may provide the researcher with a better

understanding of the students' particular writing strengths and weaknesses and thus be of greater analytic utility than either holistic or primary trait scoring (Wolcott & Legg, 1998).

#### **d) Revision Analysis**

The most common method for analyzing change across drafts has been to conduct revision analysis – that is, to count the number (and/or type) of changes the writer makes from one drafts to the next (e.g. Tuzi, 2004; Hyland, 2000; Tsui & Ng, 2000; Berg, 1999; Conrad & Goldstein, 1999; Paulus, 1999; Ferris, 1997). Both L1 and L2 researchers have relied on this type of analysis in their studies as textual changes are relatively easy to observe and quantify. In addition, it has been argued that as revision entails noticing a discrepancy between the “intended text” and the “instantiated text” (Allal & Chanquoy, 2004), revisions may be seen as evidence of change or destabilization in the learner's interlanguage – or at least greater engagement in the writing process.

Drawing upon the latter argument, Stanley (1992) used a simple frequency count of revisions to compare the effectiveness of groups trained and untrained in methods of peer feedback. In her study, thirty-one freshman ESL students were split into two groups: Group 1 received extensive coaching (seven hours over four weeks) on how to provide peer feedback; Group 2 received only an hour of training. Stanley (1992) then compared the final drafts written after the peer review session, not to “establish relative quality of the writings, but to determine the extent to which students responded to their peers' efforts by making changes in their work” (p. 233). Although no statistics were provided, the researcher claimed that essays produced by students in

the trained group contained more revisions (and thus reflected a greater level of engagement) than it did essays from the untrained group.

Other studies have taken the additional step of categorizing the revisions, typically into “surface changes” and “text-based changes” (Faigley & Witte, 1981). The former are concerned with spelling, grammar and meaning-preserving changes (such as word order), while the latter are defined as those which affect the content of the essay. From their study, Faigley and Witte’s claimed that inexperienced writers tend to make only surface-level changes to their writing, while more advanced writers make a greater number of text-based changes. If a researcher is interested in gaining a rough estimate of the relative effectiveness of various types of feedback, revision analyses may in fact be appropriate.

However, it needs to be kept in mind that a greater number of changes is not necessarily meant a writer has produced a better quality draft: A student can make small changes (too either meaning or form) without significantly impacting the quality of the text. Likewise, he or she can make a few more substantial changes which lead to a significantly better piece of writing. Thus, even when frequency counts are supplemented with categorization schemes which take into account the type and relative success of the revision, researchers need to be extremely cautious when using this procedure to make claims about the relationship between feedback and L2 writing (Faigley & Witte, 1981).

#### **e) Error Analysis**

Before any further discussion on error analysis (EA) as a form of assessment, it is indeed important to differentiate between errors and mistakes. In noticing something

wrong with a piece of written work, there is a need to decide if it is an error or a mistake. According to Klassen, 1991, “error” refers to form of structure that a native speaker deems unacceptable because of the inappropriate use. It also refers to the use of a linguistic item in a way which a fluent or native speaker of the language regards as showing faulty or incomplete learning (Richards, 1989), and also the result of not having the appropriate knowledge or having some false knowledge (Corder, 1981). So, if students are familiar with a language item but cannot correct themselves, this is an “error”. On the same note, Byrne (1993) states that learners make errors when they try to do something with the language which they are not yet able to do, like for instance, they often make false generalisations or they transfer from the mother tongue.

On the other hand, “mistakes...are slips of some kind” committed through carelessness or something that is already learned by students but temporarily forgotten (Byrne, 1993, p. 123), and also the lack of processing ability which is the ability to perform up to one’s competence level (Corder, 1981). Edge (1989) points out mistakes are caused by: (1) the influence of the first language; (2) misunderstanding a rule; (3) a decision to communicate as best one can; (4) lack of concentration, and (5) by a mixture of these and other factors. If learners make a mistake which can be self-corrected, this is called a “slip”.

According to Teh (1993), error analysis (EA) is the identification, description and explanation of errors either in its spoken or written form. There are five stages in EA. The first stage is errors identification and this is followed by errors classification (initial analysis and description of the errors), where errors are classified according to categories or sub-categories like the following: semantic errors (wrong words, wrong

form, poor choice of words, slang or colloquialism), and syntactic errors (tense, preposition, article, spelling, word order, subject-verb agreement).

Errors can also be classified as global errors or local errors. Then, an explanation to errors is provided as to why the errors have been made like mother tongue interferences, loan words, overgeneralization of rules, inherent difficulties of the target language and medium transfer, to name a few. Finally, the “errors are evaluated to determine how much they deviate from the target language norm, to what extent they affect communication and which method of correction can be most effectively meted out” (Teh, 1993, p. 54).

Research done on error analysis has provided empirical evidence pointing to emphasis on learners’ errors as an effective means of improving grammatical accuracy (White et al, 1991; Carroll & Swain, 1993). By analyzing the types of errors that students make, through observing, categorizing, and analyzing writing errors, will reveal which item has been incorrectly learnt by the students. Error analysis shed light on the manner in which students internalize the rules of the target language and develop as writers. This insight into language learning problems is useful to the teacher because it provides information on common trouble-spots in language learning that can be corrected through the appropriate remedial measures in the form of teacher’s feedback or preparation of effective teaching materials. According to Selinker (1992), errors are indispensable to learners since the making of errors can be regarded as “a device the learner uses in order to learn” (p. 150).

Xie and Jiang (2008) support error analysis as a critical tool in helping ESL and EFL learners to improve. They propose that upon identification of errors, teachers should

employ different and flexible error treatment strategies in accordance with the teaching objectives, students' linguistic competence, their affective factors and the effectiveness of the error correction. However, they also agree that high dependency on error analysis will make teacher become so preoccupied with noticing errors that the learners correct usage of second language will go unnoticed.

Corder (1973) writes that a teacher is the one who benefits the most from error analysis. Teachers who can analyse and treat errors effectively are better equipped to help their students become more aware of their errors, thus students will be able to minimize or overcome their learning problems. Error analysis with frequency counts of types of errors provides feedback and tells the teacher the effectiveness of his teaching materials and techniques, and help to discover which part of the language is not adequately learnt.

In error analysis, only those errors that alter the meaning of the text are considered important (Klassen, 1991). The decision of whether or not to correct a student's error depends on how seriously the error alters the meaning of what is being written and on the possibility that the student may self-correct the error. Teachers can categorize writing errors and examine the list to discover the patterns of errors. This information is then used to make instructional decisions.

Based on the five major techniques in assessing writing, it is rather apparent that regardless of the techniques, the goal is to provide an accurate measurement of L2 learner's performance in writing. While the techniques have their own strengths and weaknesses, what actually happens in the minds of the raters or teachers when assessing their students' writing remains a pertinent area to be investigated.

### **2.2.2 Overview of Assessment in Second Language Writing**

Due to the complexity of measuring writing, many measures exist for assessing the quality of a text, each with its own strengths and weaknesses. Writing is enormously complicated, involving not only grammatical accuracy and skills in organizing information but also a knowledge of what a particular audience expects of a writer engaged in writing a particular genre. Feedback may impact these skills in subtle ways that are not immediately measurable by any of the assessment procedures discussed above. Even when no one-to-one relationship can be found between a particular pedagogical intervention (such as feedback) and change – however defined – on the students' drafts, this does not mean that learning have not occurred (Lee, 2008).

The reversed is also true: change on a draft, for example in the form of a higher holistic score or fewer grammatical errors, does not mean that the writer has 'learned' more about the writing process. Changes can be made without understanding the reasons for making them or the ability to make them in the future on new pieces of writing. Given the evidence to date, it seems that no theory has been developed that helps explain which measures are the most appropriate for a particular set of learners, tasks and research goals. Published studies in English investigating L2 writing, known to the researcher, may have indeed employed multiple measures.

However, the majority have relied on indirect measures (such as revision analyses and interviews), which, while valuable, do not speak directly to the quality of the essay. To assess more directly the quality of a student's piece of writing, both analytic scales (which give more-in-depth diagnostic feedback on multiple aspects of the students' writing) and multiple objective measures (which can provide more detailed

information on problematic aspects of the students' grammar and lexis) are clearly needed. The acquisition of the writing skill in a second language is not easy. It is not even easy for the assessors or raters as they have to capitalize on their writing knowledge and experience in marking and assessing to be able to evaluate students' writing effectively.

### **2.2.3 Teacher Feedback in Second Language Writing**

The classroom is the context for literacy learning and it is necessary for teachers to not only examine their teaching and assessing behaviours, but also ask themselves if they implicitly allow their students to be who they explicitly encourage them to be (McMillan, 2007). There may be numerous channels by which teachers respond to students' ideas, but written feedback seems to be the most common teacher response to student work. In fact, teacher's written feedback is an essential aspect in any English language writing course. Teachers were often encouraged to adopt a process-oriented pedagogy with students ideally drafting, editing, revising, and redrafting their work. In process instruction, writing development involves a number of skills including generating ideas, organizational schemes, and goals; transforming ideas into language and its orthographic representation; and rewriting text to improve it (Berninger, Fuller, & Whitaker, 1996).

In studies that have examined it, feedback is associated clearly with writing improvement, especially for older students, and those students much appreciate it (Cheng et al., 2004; Ferris, 1997). According to Graves (1983), teacher feedback and the opportunity to revise written work based on this feedback are keys to students' development as writers. The way a teacher provides feedback will have direct impact on whether students become successful or unsuccessful writers. Sommers (1982)

believes that a teacher's written feedback could cause a student to revise his or her written work, resulting in the student producing a different piece of writing in the next draft.

Apart from that, novice writers need guidance to evaluate, modify, or restructure their ideas and to add and delete content to improve their writing (Keppner, 1991; Olson & Raffeld, 1987). On the same note, Ferris (1977) concurs that teacher feedback enables students to expand and shape their ideas over subsequent drafts of their work. Thus, with teacher assistance and feedback, students become better writers by gradually appropriating the skills necessary to critically view and revise their own work (Zellermayer, 1989).

Studies done on writing suggest that feedback plays a central role in increasing the learner's achievement. Learners of writing need to know when they are performing well and when they are not (Zellermayer, 1989). The more information learners have about their writing, the better they understand how to perform in a better way (Cardelle & Corno, 1981). Learners of writing need feedback, not only to monitor their own progress, but also to take other's view and adapt a message to it (Flower, 1979). An additional effect of corrective feedback may be the enhancement of learners' metalinguistic awareness (Swain, 1995), an important step in their appropriation of the written system.

According to Nelson and Schunn (2009), although providing feedback is commonly practiced in education, there is no general agreement regarding what type of feedback is most helpful and why it is helpful. Their study examined the relationship between various types of feedback, potential internal mediators, and the likelihood of

implementing feedback. Five main predictions were developed from the feedback literature in writing, specifically regarding feedback features (summarization, identifying problems, providing solutions, localization, explanations, scope, praise, and mitigating language) as they relate to potential causal mediators of problem or solution understanding and problem or solution agreement, leading to the final outcome of feedback implementation (Nelson and Schunn, 2009).

Teacher feedback, particularly written teacher feedback, also has a number of characteristics of interest for ongoing research into writing development. It is a key site of knowledge construction between the expert and novice, a time when teachers are most likely to provide explicit, form-focused, and individualized instruction to student writers. Teacher feedback brings into focus the language choices writers need to make to convey their ideas and often reinforces instructional points discussed in whole-class settings (Orellana, 1995; Wollman-Bonilla, 2000).

#### **(a) Research on Written Feedback**

The issue of responding to students' work has been reviewed widely (e.g. Bitchener, Young & Cameron, 2005; Ferris et al., 1997; Moxley, 1989; Straub 1997; Zamel, 1985). Patthey-Chavez and Clare (1996) investigated longitudinal development in the writings of five bilingual fourth graders and found a clear teacher influence on that development but did not pay attention to written teacher feedback between drafts. Orellana (1995) probed complex relations between teacher attitudes and the development of student writing in two very different inner-city classrooms but again did not focus on written teacher feedback between drafts. In addition, Guénette (2007) highlighted the great importance of teacher's written feedback on the student improvement.

Most research on written teacher feedback has focused on college students in general and college students who are second-language learners specifically (Baba, 2009; Ferris, 1997; Keppner, 1991; Olson & Raffeld, 1987; Zamel, 1985). For these students, researchers generally have found that teacher feedback about content (i.e., comments that encourage students to add and delete content and/ or restructure content) as opposed to teacher feedback about surface features (i.e., word choice, spelling, grammar, and punctuation) during the revision process is associated with higher-quality revisions (Keppner, 1991; Olson & Raffeld, 1987). Ferris (1997) similarly found that certain types of written comments appeared to lead to more successful revisions for college students who were English-language learners. For example, teacher requests for information, summary comments on grammar, and text-specific comments on earlier drafts of student work appeared to lead to more successful revisions, whereas statements that provided information to students and less specific comments were less successful.

### **(b) Types of Written Teacher Feedback**

Teachers usually have the choice to give both instructional and evaluative feedback on students' written work. However, the most common choice of teacher's feedback include descriptive written commentary, comments for revising major content and organization of the essay, and prose-editing responses on major editorial weaknesses (William & Robert, 1998). Basically, there are two common types of feedback: feedback on form and feedback on content. In feedback on **form**, the teacher would tend to correct surface errors pertaining to language use. On the other hand, in feedback on **content**, the teacher's focus would be in getting the students to put their

thoughts clearly in writing, so that the message can be clearly understood by the reader.

When focusing on form, teachers would typically be looking at the students' writing at the sentence level. Williams (2003) identifies three types of teacher feedback which focuses on form. The first type is where the teacher overtly marks and corrects students' errors. Another type is where the teacher indicates the place and type of error. This sometimes involves the use of error-indicating codes such as the use of 'SVA' for a subject-verb-agreement error. The last type is where the teacher merely underlines specific places at the sentence to indicate the presence of errors. Ferris et al. (1997) also discuss the form or linguistic features (or form) of teacher comments, and they suggest three variables: (1) its syntactic form (question, statement, or imperative), (2) the presence or absence of hedges in the comment, and (3) its specificity (text-specific or generic).

Feedback on content primarily involves the teacher making comments on the clarity of the message or on the cohesiveness of the text. A frequent response was to reword the student's text when it was not clear (Fregeau, 1999). Another would be to write comments in the margins or at the end of the students' written work (Bardine, 2000). Lee (2008) investigates the reactions of students in two Hong Kong secondary classrooms to their teachers' feedback focusing particularly on the factors that might have influenced their reactions. Student data from questionnaires, checklists and protocols were triangulated with teacher data from interviews, classroom observations and previous feedback analysis to situate student reactions in their specific contexts. The results show that students, irrespective of proficiency level, wanted more written comments from teachers. The students of lower proficiency were less interested in

error than those of higher proficiency, though both groups preferred more explicit error feedback from teachers. The students who did not understand the entire teacher feedback could be due to its illegibility, apart from other plausible factors not explored in the study.

Other researchers had different ways of categorising feedback or comments. For instance, Caulk (1994) in his study suggests that comments can be divided into six categories: (1) **form** – suggestions to change introductions, conclusion, and paragraphs so they would be closer to typical English form (for instance, making the general point clear in the introduction, not bringing up new arguments in the conclusion, etc.); (2) **reorganization** – suggestions to change the order of words, phrases, sentences, and paragraphs for reasons not due to form; (3) **more information** – suggestions to write more detailed information about one aspect of the paper; (4) **write less** – suggestions to write less information about one aspect of the paper; (5) **clarity** – statements that the meaning of a particular sentence, point, or paragraph is unclear to the reader and suggesting it to be made clearer; and (6) **style** – suggestions that the style of a particular sentence or passage is not the most effective for that particular writing task.

In another instance, Olson and Raffeld (1987) categorize the type of feedback students received on each draft as **surface level**, **clarification level**, or **content level**. Surface-level feedback included all edits and comments students received that pertained to mechanics, usage, grammar, spelling, sentence structure, or format. Clarification-level feedback was defined as teachers' directions to students to clarify or elaborate on specific words. Content feedback was defined as teachers' comments pertaining to the concepts in and structure of students' writing. This type of feedback

included comments to delete, reorganize, or add information, as well as questions intended to challenge students' thinking. This category of feedback also could include genre-specific comments or edits.

### **(c) Error Corrections (EC) as Feedback in Written Work**









According to Teh (1993), there is no single method of dealing with errors made by students. Some of the common practices of teachers are to mark every error, provide the correct answer for errors made, mark the first and only draft or work written by students, make general comments, make students rewrite the corrected version several times over and view errors as signs of failure. Josephson (1989) in his study has advocated some of the method for error corrections: the use of peer marking/editing, selective marking, code correction, effective and specific comments, checklist of limited common errors, different coloured inks, discussion of errors on tape and direct versus discovery-type of marking.

In supporting peer marking/editing method for EC, Byrne (1993) proposes that students be allowed to work at these mistakes/errors themselves with the help of their peers. Peer-marking/editing is especially useful in the first draft of their written work. Here students are given the responsibility to edit each other's work individually or in a group before handing in the final draft to the teacher. Besides being fun for students to be allowed to correct and learn from errors other than their own, it also reduces the need for too many red markings from the teacher.

Klassen (1991), however, opines that not every error should be corrected and believes that gravity of errors should determine which correction is necessary. She suggests that teachers should focus on marking global errors in the first draft of their student's

written work and then local errors in the second draft. The teacher could provide clues and codes in the form of abbreviations, symbols, arrows, circles, lines and explicit marginal comments (see Table 2.1), instead of providing the correct answers every time a student makes an error. The practical reason for this is students will continue to make that error unless they recognize the type of error they are making.

---

	Serious error		Minor error
	Written in margin – multiple errors		Omission
	Error in paragraphing		Repetition (Circle the repeated words/structure)
	Merit structure/ vocabulary)(written above the word/structure)		Paragraph

---

Table 2.1 Samples of the symbols used in the marking of students' writing

Haycraft (1993) suggests that the most comprehensive method of correcting compositions is to have an analytical approach where mistakes are divided into categories and abbreviations. Haycraft (1993) also notes that teachers should underline errors in the text and put appropriate symbols in the margin to indicate the type of error. Thus, at the same time students would be able to see what they originally wrote. At the end of the composition, the errors are added up under each category and listed. Through this approach, the students will become more conscious of their common errors and may be careful in writing subsequent compositions. Byrne (1993) also finds correcting all errors in students work time-consuming for teacher and discouraging for the students. Byrne (1993) believes that teachers should:

...not attempt to correct all mistakes in a piece of writing but only those in certain areas, such as tense or articles, either because this is where students particularly need help or because you (teachers) have decided to *focus* attention on these for a while.(p.125)

By focussing on a few errors each time, students are less likely to become demotivated in learning English (Ellis and Sinclair, 1989).

As opposed to the advantages of EC, numerous studies have revealed that grammar correction to second language writing students is actually discouraging to many students, and even harmful to their writing ability (Semke, 1984; Sheppard, 1992; Truscott, 1996). Truscott (1996) reasons out that writing class grammar feedback does not work because it treats only the surface appearance of grammar and not with the way language develops. Indeed, those who support this opinion believe that learning of linguistic items does not occur in a linear fashion, considering that learning grammar in a second language is a complex and gradual process. Thus, for grammatical correction to work, the correction must be precisely tied into the correct levels of this process.

Research has also shown that corrections made by second language writing teachers are frequently arbitrary, not consistent, and greatly dependent upon the amount of time the teacher has with L2 students (Zamel, 1995). Apart from that, students often find teachers' remarks vague, confusing, and contradictory, and they also feel that teachers do not provide sufficient grammatical explanations about their writing mistakes (Cohen, 1987). According to Cohen (1987), generally students only make a mental note of the corrections they have understood, and usually would not incorporate these corrections into their work if they have to rewrite their papers.

#### **(d) Impact of Teachers' Feedback**

There are positive and negative effect derived from the form-focused feedback. On a positive note, form-focused feedback was found to significantly improve a student's

grammatical score on subsequent rewrites produced by the students (Fathman & Whalley, 1990; Frodesen, 2001, Hyland, 1998; Inbar-Lourie, 2008). A better understanding of feedback could turn the situation into ground for teachers and students to engage in a dialogue that seeks to “understand the different assumptions regarding texts, writers, audiences and knowledge that writers can have” (MacKay, 1993, p. 78).

However, on the other hand, such feedback could be a source of miscommunication and misunderstanding (Hyland, 1998) and thus, would not be able to help develop a student’s writing skills. In theory, constructive feedback provides learners with means of understanding their readers' needs, but in practice, most learners do not receive such feedback (Zellermayer, 1989).

Research findings from second language classrooms indicate that writing teachers are mainly concerned with responding to surface level features and mechanical errors. Such feedback is likely to be ineffective, confusing, inconsistent and contradictory (Zamel, 1985; Robb, Ross & Shortreed, 1986; Cohen, 1987). A number of writing instruction theorists (Semke, 1984; Sommers, 1982) have pointed out that feedback provided and processed inadequately may inhibit the learner's motivation for writing.

Williams (2003) points out that when students merely copy teacher corrections, they become passive and are not able to recognise, nor correct errors on their own. Allwright (1975) in his study observes that error treatment in the classroom is imprecise, inconsistent, and ambiguous. He discovered that feedback was confusing to learners in that learners often received contradictory signals simultaneously with respect to the content and the form of their utterance.

On a similar note, Cohen and Cavalcanti (1990) in their review conclude that teacher feedback was “unclear, inaccurate, and unbalanced, “causing students to either ignore the feedback, feel hostility towards it, or make only minor sentence-level changes in their draft” (p. 155). Fregeau (1999) in her study discovered that students’ writing skills were not enhanced even when a teacher merely indicated the presence or types of errors without correcting them. This was because the students did not understand why the errors were indicated and simply guessed what corrections were to be made.

Where content is addressed, a problem may occur if a teacher were to impose his or her values and understanding on the students’ writing (Li, 1995). A teacher in an attempt to correct the content will impose his or her perceived ‘ideal text’ on what the writer really wants to say. This effort on the teacher’s part is fairly time consuming and also ‘risky’ activity. According to Berg (2003), there is a possibility of misinterpretation and thereby commenting or correcting errors in a way that changes the meaning that the writer intended. Sometimes, problems could also arise when teachers write vague comments in the margins or at the end of a student’s written work. Vague comments such as “Very good”, “Nice, but redo” often confuse students rather than help them revise their writing (Fregeau, 1999).

These comments do not give any specific suggestions or directions to students about how to proceed in their revision of text. Leki (1990) found that students’ responses to teachers’ feedback on content is also quite discouraging. It was found that some students do not read the comments at all. Some students may read the comments but do not understand what the comments mean, whilst others may read and understand the comments but do not know how to respond to the comments.

So far, the literature has painted quite a disheartening picture on the impact of teachers' feedback on students' writing. So are we wasting our time marking students' writing? What we need to acknowledge is the fact that not all our written feedback can help our students improve their writing. Thus, we need to ensure that the time taken to mark our students' writing is time well spent.

#### **(e) Teachers' Role in Providing Feedback**

The teacher takes on an active role in the editing process in process writing but there has been considerable debate on what type of feedback should be given (Hattie & Helen Timperley, 2007). According to Radecki and Swales (1988), "there is an emerging consensus among certain L1 and L2 writing researchers that an instructor's time is better spent in attending to textual meaning, rather than to grammatical errors, as a means of aiding students in reshaping their writing" (p. 72). Ferris and Hedgcock (1998) believe that such grammatical corrections are necessary but should not be given on first drafts of multiple draft essays. Comments on earlier drafts should, according to the authors, focus on organization and content rather than grammar. In contrast, Fathman and Whalley (1990) found that "grammar and content feedback, whether given alone or simultaneously, positively affect rewriting" (p. 185).

In responsive teaching, the student acts and the teacher reacts. The range of reaction is extensive and diverse because individual teacher is responding or giving feedback to an individual student, and the students in turn is passing through an ever-changing process of discovery through writing (Murray, 1993). Citing researches done by Lynch and Klemans (1978), Reed and Burton (1981), and Burkland and Grimm (1984) on writing, Radecki and Swales (1988) note that the students stated a

preference for their instructors to edit grammatical and other mechanical errors as they found these comments to be the most helpful.

Teachers therefore, need to come up with an effective method of feedback to help students improve their writing. Teacher's feedback on the idea development, rhetorical structure, style and organization can vary considerable. According to Straub (1997), teachers' comments can either be directive or facilitative, general or specific. However, these responses sometimes do not provide students with a clear understanding of what they are supposed to do when they are revising their text (Zamel, 1985). In order to ensure affective feedback is rendered to the student, several issues need to be addressed.

First and foremost, before a teacher can determine how to provide feedback, he or she needs to have a clear understanding of the objectives of the particular writing course taught. Teachers' feedback which does not reflect the writing course objectives may result in students not knowing how to respond appropriately (Williams, 2003).

Another issue that needs to be looked into is on the quantity of errors that a teacher can identify or the quantity of written comments that a teacher can make. Dohrer (1991) feels that teachers should not make too many comments on students' writing. This is because students feel frustrated when they see too many written comments on their paper. Omaggio (1986) believes that it is important for teachers to determine what errors to respond to, as identifying and getting students to correct all the errors can destroy the students' self-esteem and confidence.

Finally, the way teachers write their comments should also be considered. Teachers' written comments need to be detailed and descriptive so that students can understand them. Writing one word comments such as "unclear" or "rewrite" should be avoided (Doher, 1991; Bardine, 2000). Omaggio (1986) suggests using leading questions, giving hints about what corrections should be made as possible avenues for teachers to help students improve their writing. Positive feedback where the teacher recognises and reinforces what the student has correctly acquired is also believed to be effective in enhancing students' writing skills (Fregeau, 1999).

#### **2.2.4 Overview of Teacher Feedback in Second Language Writing**

In general, learners need to receive feedback in different ways for different reasons. In fact, ESL teachers should provide meaningful and constructive feedback. Their feedback should deal with all aspects of the learner's writing. This feedback should encourage the learner to think about, write and rewrite multiple drafts of each topic. Teacher's feedback should not focus on direct correction of errors only, but lead the learner to write clearly, unambiguously, logically, and coherently.

Based on studies done on feedback, it is very clear that quality feedback by teachers can stimulate students' awareness of their writing as they develop the skills to draft, revise and edit their own work. Lastly, there is a need to recognize that related factors, like student beliefs, reactions and expectations about the teacher's feedback, constitute a major determinant that influences achievement in writing (Myles, 2004).

#### **2.2.5 Evaluating Writing**

Evaluation is an important element in the teaching and learning process (Chitravelu et al., 1995). The purpose of an evaluation affects the evaluation procedures and how the

results are analysed and used. Evaluation in writing is important to motivate students to write better through showing the progress they have made by appreciating their writing. This kind of evaluation is also to point out the students' strengths and weaknesses and to help them to improve their writing.

In evaluating writing, a teacher plays a multiple role as a judge, reader, adviser, and also as a resource person (Chitravelu et al., 1995). In testing, the teacher plays the role of a judge. When the teacher plays the role of a reader, he/she needs to be a real reader, informing the students his/her dislikes, disagreement and vagueness. When the teacher plays the role of an adviser, the teacher gives ideas to improve on the writing, for instance, to add interest to the story or add more descriptive words. As a resource person, a teacher can help students with new vocabulary, guide them in making references, and be readily available in case they need to discuss any related matter.

In evaluating the written products, usually a fixed set of criteria such as performance bands is used (Chitravelu et al., 1995). This performance band helps the teacher or examiner evaluate his/her student's performance in achieving the objective of the course. Evaluation is a continuous process and its aim is to help individuals progress, to make comparison with his previous performance. It is presumed that checking students' writing is a tedious task that demands ample amount of time. No matter how long it takes or how big the number of students, teachers of English consider it important to spend time and energy checking students' writing with great care.

The evaluation of the students' writing is normally subjective. There is a need to construct a rubric, a kind of scoring grid that elaborates the elements of writing that are to be evaluated. This rubric should outline the weight of grammar and mechanics

in relationship to content and ideas, as well as other feature of writing that teachers find relevant and important. There are three general types of rubrics that can be used in evaluation; the non-weighted rubric, the weighted rubric and the holistic rubric.

The **non-weighted rubric** provides descriptions of writing quality by level across other writing criteria. For instance, the content, organization and grammar of the students' writing would be evaluated and given written comments based on the quality of writing - excellent, good and poor.

The **weighted rubric** is similar to the unweighted one, but it breaks the writing skills into categories and sub-categories. A specific point is assigned to each category and sub-category, taking into consideration the following elements:

- it has a clear introduction
- it has separate paragraphs
- it has a conclusion
- it uses transitions to join paragraphs
- it uses transitions when needed within paragraphs

The **holistic rubric** describes in general terms the qualities of excellent, good, fair, and unsatisfactory assignments (see Appendix 9 - A Sample of Marking Scheme Used in a School). These descriptions can be tied to grades. For instance, a grade 'B' (a competent writer) paper shows:

- Very competent writing
- Minor errors present
- Vocabulary is precise and fairly wide
- Some variety of sentence structures

- Accurate Punctuation
- Accurate spelling
- Organised and suitably linked paragraphs
- Relevant content

### **2.2.6 Theoretical Framework of the Study**

This theoretical framework is constructed to provide a comprehensive representation of relationships between the different types of assessment and evaluation knowledge (as discussed in sub-section 2.2.1), teacher feedback (as discussed in sub-section 2.2.3), error corrections (EC) in written work (as discussed in sub-section 2.2.4) and the teachers' skill in evaluating students' written work (as discussed in sub-section 2.2.5).

The factors that contribute to good writing are derived from such discussion, and are further categorised into five divisions that include grammar, mechanics, content, organisation and vocabulary. All these form the knowledge states which are required, along with the assessment and evaluation knowledge, by teachers/raters in assessing and giving feedback on students' writing. The Theoretical framework of the relationship between assessment and evaluation knowledge, and teacher skill in assessing and giving feedback on students' written work is shown in figure 2.1 below.

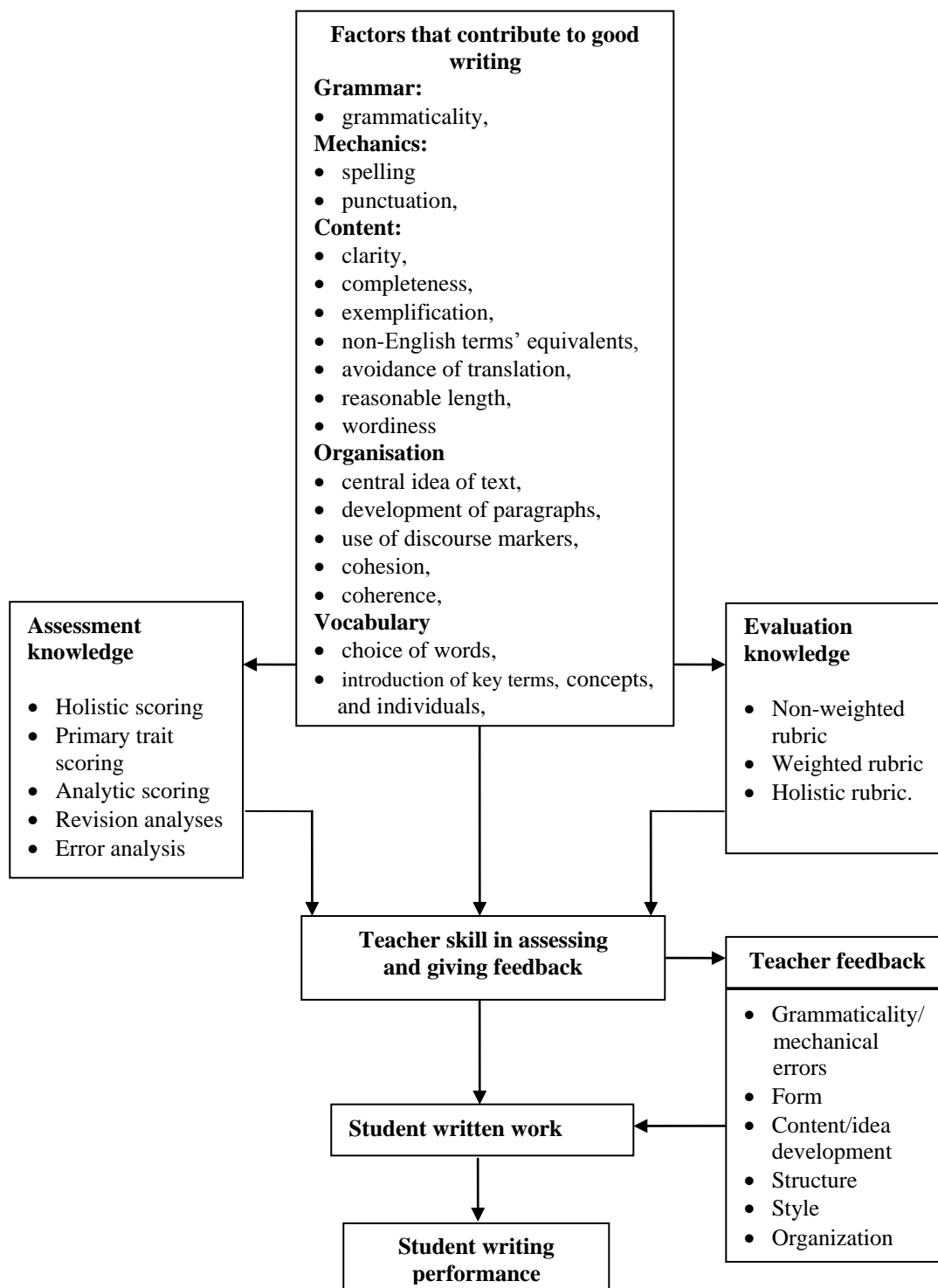


Figure 2.1 Theoretical framework of the relationship between assessment and evaluation knowledge, and teacher skill in assessing and giving feedback on students' written work

## **2.3 Literature Review Pertaining to the Variables of the Study**

This section presents and discusses the literature review on the variables of the study. It starts with a brief review of the cognitive analysis of problem-solving which leads to the focus on knowledge states and conceptual operators. Reviews of literatures over knowledge states in writing lead to the classification of various elements that contribute to clear writing to be used in the data analysis of this study. On the other hand, the literature reviews on conceptual operators are largely based on ones used in the medical domains (Hassebrock & Prietula, 1992) in the analysis of medical problem-solving protocols.

### **2.3.1 Cognitive Analysis of Problem-solving in Assessing Writing**

In order to understand how the expert and novice ESL raters differ in assessing and giving feedback on students' written work, in terms of their deep structure of knowledge in a manner similar to the study of expertise in the medical domains (Hassebrock & Prietula, 1992), a detailed cognitive analysis of the participants' verbal protocols in assessing writing task is needed to examine the assessing mechanism used. In this present study, Cognitive Task Analysis (CTA) was used to tap the knowledge raters use when they assess writing. It would illicit behaviours that manifest this knowledge.

In conducting a protocol analysis, one develops a coding scheme which provides both an explicit framework for documenting behavioural events and a guide for structuring the analysis. As stated by Hassebrock and Prietula (1992), the development and application of a coding scheme depends primarily on:

- the nature of the task (a theory of the domain)
- the theoretical constructs of interest (a theory of reasoning in the domain)

As cognitive processes are only indirectly and partially represented in verbal reports, it is necessary to analyse protocols by means of a coding scheme that will guide the researcher's inferences in a principled, theory-based manner. A coding scheme should be a theoretically grounded model of the cognitive processes and the types of information involved in the activity under study are not a mere list of strategies. Even if the verbal report data are used only as supportive data rather than as the primary data set of a study, they have to be analysed according to an appropriate coding scheme if they are to be taken seriously (Bracewell, 1994; Greene and Higgins, 1994).

According to Hassebrock and Prietula (1992), the first step in analyzing a protocol is to break down the transcript into short segments or phrases. Each segment reflects an "assertion" made by a subject at a given instance. The boundaries of segments can often be determined by noting syntactic cues such as clauses, sentences and pauses in the protocol. This step in analyzing the protocol yields a topic representation in which each segment addresses a particular instance of reasoning behaviour on the task. The topic representations are then coded depending upon the analyst's decision as to how a particular topic segment maps onto the constructs or categories of the coding scheme.

A given topic segment can consist of knowledge states which are units of information or knowledge retrieved from memory and used by an individual. Second, these knowledge states are associated with conceptual operators which are inferential processes that produce new or modify existing knowledge states. The general goal of the analysis is to trace the sequence and pattern of knowledge states and conceptual operations used in response to a problem-solving task.

Newell and Simon (1972), who propose a scheme to code topic representations, suggest that protocols consist of two types of semantic elements: knowledge state and conceptual operation. However, Hassebrock and Prietula (1992) in the analysis of medical problem-solving protocols have included the third element that is the lines of reasoning. The third element is known as the line of reasoning which can be viewed as a problem-specific representation of problem-solving knowledge. In their study, Hassebrock and Prietula (1992) view the line of reasoning as a means of representing physician's use of more global diagnostic strategies and problem-solving representations, based on the analysis of the knowledge states and conceptual operators used by individuals during the think-aloud process of the problem-solving task.

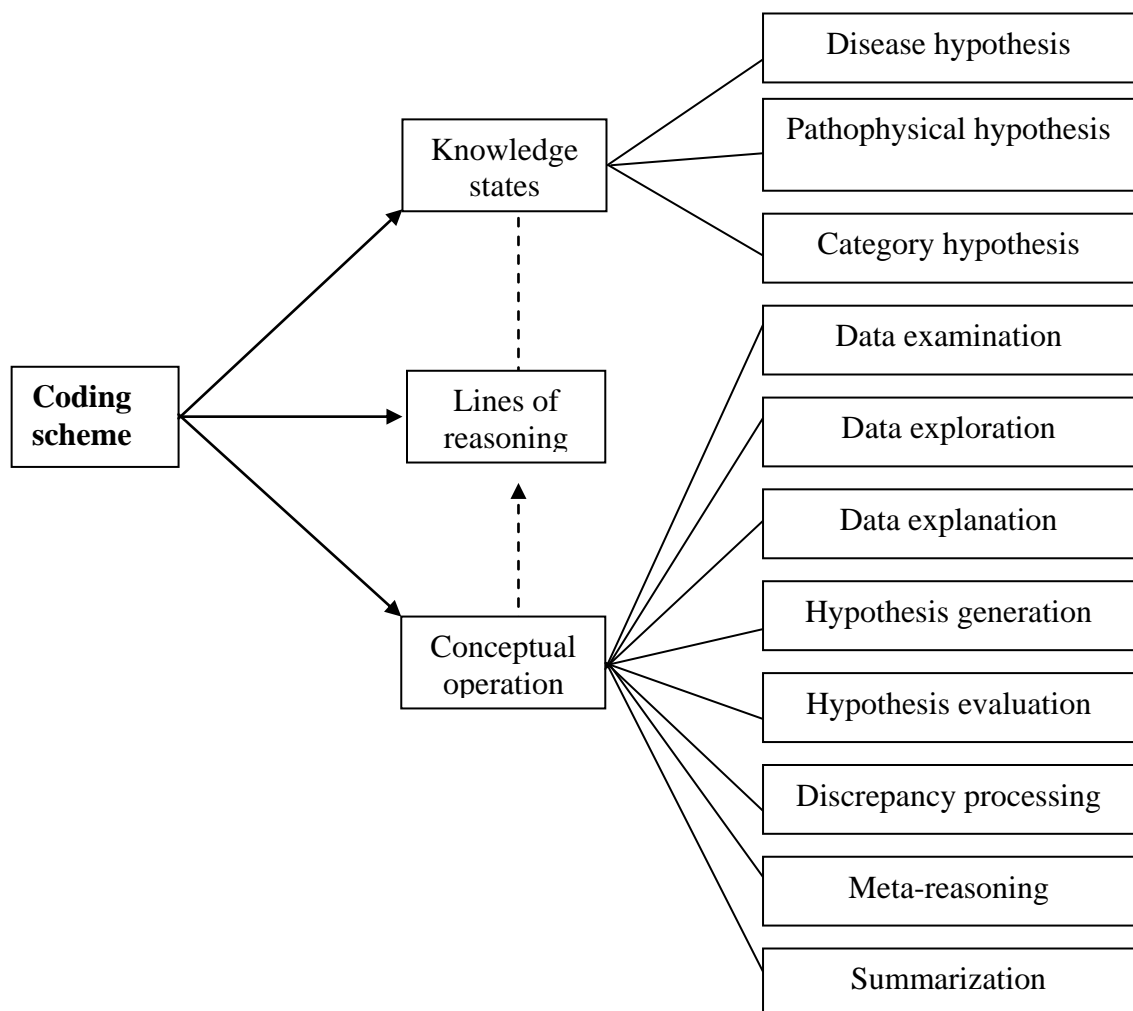


Figure 2.2 Basic elements of the coding system for the analysis of medical problem-solving protocols (Hassebrock & Prietula, 1992, p. 622)

The coding scheme presented in Figure 2.2 above has been constructed by Hassebrock and Prietula (1992) for analyzing verbal protocols from diagnosticians of congenital heart disease. The scheme serves as a guide for structuring the analysis and is used to identify representations of reasoning behaviour that differ in specificity and conceptual focus.

**(a) Knowledge States**

As shown in Figure 2.2, the first protocol representation in the coding scheme is the knowledge state. It identifies the units of medical knowledge used by the diagnosticians of congenital heart disease, namely, disease hypothesis, pathophysical hypothesis and category hypothesis. These knowledge states are associated with conceptual operators which are inferential processes that produce new or modify existing knowledge states that help in the analysis of the related problem-solving task.

**(b) Conceptual Operation**

The second protocol representation in the coding scheme is the conceptual operation. It is an inferred cognitive process which modifies (adds, eliminates) existing or currently active knowledge states and produces new, active knowledge states (Newell & Simon, 1972). In response to a specific data cue, a given segment of verbal protocol will constitute one or more knowledge states and a conceptual operation that produces the associated knowledge state or states. The eight types of conceptual operations and their specific operators used by Hassebrock and Prietula (1992) for analyzing verbal protocols from diagnosticians of congenital heart disease are shown in Table 2.2. They are data examination, data exploration, data explanation, hypothesis generation, hypothesis evaluation, discrepancy processing, meta-reasoning, and summarization. These conceptual operations are used to characterize distinct segments of a

physician's problem-solving behaviour. Each basic conceptual operation is further analysed to give a more detailed representation of knowledge and reasoning behaviour required by the task.

<u>Conceptual Operations</u>	<u>Conceptual Operators</u>	
Data examination	a) Read b) Identify c) Examine	1) Compare to norm 2) Compare-to-expected 3) Determine-severity 4) Compare-over-time 5) Identify-location 6) Compare-multiple
Data exploration	a) Examine b) Scan c) Search d) Elaborate e) Integrate f) Note-absent-data	1) Compare to norm 2) Compare-to-expected 3) Determine-severity 4) Compare-over-time 5) Identify-location 6) Compare-multiple
Data explanation	a) Infer-pathophysiological-cause b) Infer-natural-history	
Hypothesis generation	a) Trigger b) Further-specification c) Association d) Causal-relationship e) Generalization	
Hypothesis evaluation	a) Confirmation b) Disconfirmation c) Discrimination d) Causal-relationship	
Discrepancy processing	a) Recognition b) Resolution	1) Ignore 2) Explain-away 3) Systems-thinking
Meta-reasoning	a) Experiential-memory b) Cue-diagnosticity c) Diagnostoc-plan d) Self-evaluation	
Summarization	a) Repeat-data b) Repeat-hypothesis	

Table 2.2 Detailed descriptions of conceptual operations (Hassebrock & Prietula, 1992)

## **1. Data-Examination**

The physician's goal in this problem-solving episode is to select and examine cues and make initial interpretations of the finding in the patient data of a case.

- (a) Read. The subject only reads a specific data and does not generate any thinking-aloud protocol.
- (b) Identify. The subject selectively identifies or repeats a specific cue or cues from that particular set of data. The subject does not provide any additional thinking-aloud protocol in response to the particular cues identified.
- (c) Examine. This operator is used to determine the significance of a cue. There are six variants for this examine operator. They are: (1) compare cue to the norm; (2) compare cue to its expected probability; (3) determine degree of severity or abnormality; (4) compare cue's value over time; (5) identify location of a cue; and (6) compare multiple cues.

## **2. Data-Exploration**

The physician's goal in this episode is to conduct a more detailed interpretation of cues previously selected from the case.

- (a) Examine. Additional cue interpretation by using any of the examine operators described above (see *examine* in data-examination).
- (b) Scan. A subject conducts a sequential review by identifying or listing a specific set of data.
- (c) Search. A subject states that a reason for reviewing or examining the previously selected cues is to ascertain whether a specific cue or finding existed or was present in the case data.

- (d) Elaborate. This operator is used to note that the description of a particular cue lacks specific information and that additional elaboration would be necessary for complete interpretation.
- (e) Integrate-symptoms. A subject uses this operator to identify a small number of data cues from different categories of patient data. This group of data is integrated for subsequent or additional interpretation.
- (f) Note absent data. This operator is used to note that a particular cue is not in the data.

### **3. Data-Explanation**

The physician's goal in this episode is to interpret the significance or role of a given data cue by explaining the underlying pathophysiological basis of a disease. It is used to count for the natural history which could lead to the development of data.

- (a) Infer-pathophysiological-cause. In using this operator, a subject goes beyond merely noting whether a cue represents an abnormal finding. It is used to infer a pathophysiological condition underlying the presence of a particular cue.
- (b) Infer-natural-history. This operator is used to explain the developmental course of a particular cue with respect to a disease or pathophysiological condition.

### **4. Hypothesis-generation**

The physician's goal in this episode is to generate diagnostic hypotheses. The specific operators provide different means of generating hypotheses in response to a particular data cue, pattern of cues or from other diagnostic hypotheses.

- (a) Trigger. This operator generates the first use of a hypothesis.
- (b) Further-specification. This operator triggers a hypothesis which is subsumed by a previous generated hypothesis.

- (c) Association. With this operator, a new hypothesis is formed through combining two or more previously triggered hypotheses.
- (d) Causal-relationship. This operator uses a previously generated hypothesis to generate another hypothesis through implied causal relationship.
- (e) Generalization. This operator is the inverse of further-specification. Sometimes it is necessary to “relax the constraint” when the generated hypothesis is too specific, and consider a more general, but related type of hypothesis.

## **5. Hypothesis-evaluation**

The goal of this episode is to evaluate a hypothesis by interpreting its ability to account for the presence or absence of data cues, data interpretations, or its relationship to other diagnostic hypothesis.

- (a) Confirmation. A cue is interpreted as being consistent with a hypothesis and provides a confirmatory value for the hypothesis.
- (b) Disconfirmation. A cue is interpreted as being inconsistent with a hypothesis and provides a confirmatory value for the hypothesis.
- (c) Discrimination. This operator is used when a specific cue or several cues are interpreted with respect to two or more diagnostic hypothesis.
- (d) Casual relationship. This operator is used when expected or implied causal relationship between two or more diagnostic hypotheses is considered as confirmatory evidence.

## **6. Discrepancy-processing**

The physician’s goal in using this operator is to identify and/or resolve discrepancies that exist between the presence of specific items of patient data and the presumed presence or absence of an underlying disease or pathophysiological condition.

- (a) Recognition. A subject states a discrepancy or describes an anomalous situation existing among one or two more data cues and one or more knowledge states.
- (b) Resolution. A recognized discrepancy may be *ignored* if the subject merely stated what cue would be expected or if a discrepancy has serious implications for evaluation of a disease or group of diseases, a subject might *explain-away* the discrepancy by questioning the reliability of the reported clinical findings

## **7. Meta-reasoning**

The physician's goal in this episode is to evaluate the reasoning process.

- (a) Experiential-memory. A subject uses this operator to recall related information and also retrieve related general knowledge.
- (b) Cue-diagnosticsity. A subject used this operator to make a general comment on the domain of cardiac entities that are plausible explanations for a case, though there is no specific hypothesis being provided.
- (c) Diagnostic-plan. A subject can also use this operator to provide specific comments on the diagnostic plan or strategy that is being adopted.
- (d) Self-evaluation. A subject uses this operation to criticize or critique the adequacy of specific diagnostics hypotheses or general diagnostic conclusions.

## **8. Summarization**

The goal of this episode is to maintain data or hypotheses in active memory. Based on the above eight types of conceptual operations and their specific operators (Hassebrock & Prietula, 1992), the same conceptual operations are adapted and used in this study to characterize distinct segments of a rater's problem-solving behaviour in assessing writing. This is further discussed in Chapter 3 under sub-section 3.7.1(c).

Each basic conceptual operation is further analysed to give a more detailed representation of knowledge and reasoning behaviour required in assessing writing.

### **(c) Lines of Reasoning**

The use of lines of reasoning (LoR) is analogous to the development of a problem representation or a mental model as stated in Section 2.3.1. A LoR is not a pre-stored knowledge unit “in the subject’s mind”; rather, it is a reasoning artifact representing a complex set of related pathways of thought bound together (Hassebrock & Prietula, 1992, p. 630). The knowledge content of a LoR, based on the analysis of the knowledge states and conceptual operators used by an individual in the problem-solving task, serves to discriminate expertise as well as individual differences among experts.

Thus, through a series of procedures, the verbal protocol analysis used in the study has identified the knowledge states and conceptual operations produced by a physician when solving a medical problem-solving case. Similarly, VPA may provide a useful method of analysing thinking to yield information about the thought processes and structures underlying the task of assessing writing (Crutcher, 1994).

## **2.4 Literature Review on Methodology**

This section discusses review of literature on cognitive task analysis, cognitive task analysis methods, verbal protocol analysis, metacognitive strategies and skills, problem solving, the mental model and procedures in analysing verbal protocol analysis which relates to the methodology of this study.

### **2.4.1 Cognitive Task Analysis**

Cognitive Task Analysis (CTA) is commonly accepted as an approach that is used to obtain data about cognitive skills that are embodied in the task. CTA uses interviewing and modelling techniques as well as experimental procedures to determine the cognitive structures (e.g. mental model and memory organization) and processes (e.g. decision making and problem-solving) which underlie the skilled job performance. It provides insight into the cognitive processes in the mind (Militello & Hoffman, 2008). The knowledge structure of the task allows us to see how knowledge is organised and concepts are interrelated.

CTA identifies skills needed for the task performance of the task. Determining the mental models in a task performance is a unique feature about CTA that is never found in any traditional methods of data analysis. As stated in a recently reported research by Hoffman, Neville and Fowlkes (2009), CTA permits thorough and explicit documentation of critical decision-making procedures that are otherwise inaccessible through surface research methods such as interviews and survey. Shao and Yu (2008) further demonstrated in their studies how CTA can be used to elicit important aspects of test item difficulties.

Apart from that, CTA also attempts to identify information-processing strategies. Usually a person doing the same problem could come out with several strategies to reach a similar solution. Redding (1995) categorises three aspects in CTA; determining of cognitive structures, recognising skills needed to perform a task and analyse the mental model. Comparisons are between the experts and those with less experience to determine the following:

- How a task is learned

- How to expedite learning
- Optimal job performance
- Required job performance skills

Such findings are especially useful to train a novice to become an expert and to improve teaching and learning process in school.

#### **2.4.2 Cognitive Task Analysis Methods**

According to Gordon (1994), some of the methods for conducting CTA include structured interviews and verbal protocol analysis. Nevertheless, CTA cannot provide all the necessary data about the mental process; we need behavioural methods to supplement data. Chipman et. al. (2000) therefore suggested the use of observation on the task performance during the verbal protocol task, where analysis is based on the observable aspects of the task. Besides that, semi-structured interview can be used to identify the abstract nature of the knowledge involved in the task, especially the types of knowledge representations being used. Yin (1994) pointed that interviews are a useful source of collecting evidence for the case study as they give the interviewees' constructions of the reality around them and may help provide important insights on how they develops their understanding and interpretation of concepts and procedures involved in problem-solving task. Patton's (2002) suggested five types of questions to aid knowledge elicitation - experience/behaviour questions, opinion/values questions, feelings question (affective questions), knowledge questions, and sensory questions (Appendix 7).

#### **2.4.3 Verbal Protocol Analysis**

Crutcher (1994) stated that verbal protocol analysis (VPA) has emerged as one of the main methods for analysing thinking. The use of protocol analysis to infer cognitive

processes is based on theoretical assumptions about human cognition that must be understood to use the method properly. Ericsson and Simon (1991) states that;

A single verbal protocol is not an island to itself, but a link in a whole chain of evidence, stretching far into the past and the future that gradually develops, moulds, and modifies our scientific theories. It needs to be processed with full attention to these linkages. (p. 280)

With increased interest in human information processing and artificial intelligence research, verbal protocol analysis have become an important tool for examining problem-solving behaviour and have been widely used in the domain of cognitive psychology such as decision making, text comprehension, the investigation of cognitive processes in writing, test taking and also test validation (Ericsson and Simon, 1991).

According to Newell and Simon (1972), the distinct advantage of using the verbal report data is that it occurred in the context of ongoing problem-solving behaviour. It has an advantage of allowing the convergence between what was said with what was actually done by the subjects. There are a lot of arguments that are against and for this so-called direct research method and thus many people still seem to have misconceptions about it. By giving the same problem to a number of subjects, the data collected has allowed Newell and Simon (1972) to see the invariant aspects of the solution process and avoid idiosyncracies in an individual's protocol.

Verbal protocols obtained through the talk or think-aloud method would be useful in a wide range of tasks and domains, including assessing writing, as an effective tool for understanding how cognitive skills and strategies change and develop (Robinson, 2001). This is because verbal protocols can provide more complete information on writing raters' problem-solving/assessing strategies. Think-aloud protocols also provide a means of assessing the mental processes of an individual (Pugalee, 2004).

There are many methods that can be used to generate verbal reports such as by asking questions, reporting mental processes, and by asking the subjects to talk-aloud or think-aloud. Although data obtained from this method is not an isomorphic mental operation, it does provide a direct overall view or at least a partial view of the strategy used in problem-solving which cannot be obtained from any other methods (Newell & Simon, 1972). This is further supported by Redding (1995) who states that the talk-aloud or think-aloud method is able to provide the most accurate records about mental processes on their nature as well as sequence.

In the **think-aloud activity**, researchers instruct their subjects to verbalise directly only the thoughts entering their attention while performing the task (Ericsson & Simon, 1991). This type of verbal reporting is concurrent with the execution of the specific task, and the subjects provide information while it is still available to them in the short term memory. Based on the literature, the researcher concludes that protocol analysis is particularly useful for investigating problem-solving and mental models.

However, verbal protocols are not immediate exposure of thought processes. They represent (a subset of) the information currently available in short-term memory rather than the processes producing the information. Cognitive processes are not directly manifest in protocols but have to be inferred, just as in the case of other types of data. Hence, the researcher believes that the main concern for researchers, in revealing the knowledge and cognitive processes used by a person while performing a task or behaviour, is how the protocol data are analysed. This methodology is often referred to as “process tracing” since the continuous stream of verbalizations provides evidence for the underlying representations of knowledge and the covert

psychological processes which contribute to a final task outcome or problem solution (Chipman, et. al.. 2000).

Verbal protocol can also be carried out either concurrently, retrospectively, or prospectively (Hoffman, 1987). We can do ***concurrent protocol*** by asking subjects to perform some tasks such as problem-solving and at the same time asking them to think aloud. It is also referred to as ***introspection*** (Ericsson & Simon, 1991). The verbal report is usually recorded on tape, later transcribed, coded from the transcript, and then analysed. However, sometimes it is difficult to conduct concurrent protocols, especially if the event takes place very fast or is too cognitively demanding, so retrospective protocol can be used as an alternative (ibid.). ***Retrospective protocol*** is sometimes done after concurrent protocol as a supplement to provide the missing information or to fill the gaps in concurrent protocol. While analysing the retrospective protocol data, caution should be taken as the subjects may reconstruct events that did not actually occur while performing the task.

Ericsson and Simon (1980, 1984) (cited in Hassebrock & Prietula, 1992, p. 616) claim that verbal reports are indeed valid as psychological data as much as other, more typical, sources of behavioural data obtained from experimentation. They assume that any verbal report has to draw upon information that has been stored in a person's memory system. Information processing theory specifies three general types of memory stores including a sensory memory (SM) of very short duration, a short-term memory (STM) with limited capacity and intermediate duration and long-term memory (LTM) with potentially permanent storage and large capacity. Consciousness or immediate awareness involves only the informational content that is currently maintained in STM. Thus, sensory and perceptual information stored in SM must be

recognized and encoded into STM and conceptual and semantic information contained in LTM must be activated or retrieved into STM before a subject could verbally report on either type of information.

The completeness or quality of a verbal report at a given time is determined by the inherent limitations of the human information processing system, as a subject's verbal protocol will contain only information stored in STM at the time of the report. Ericsson and Simon (1980, 1984) (as cited in Hassebrock & Prietula, 1992) opines that "concurrent verbalization provides the most complete report since information is verbalized as it is being attended to while performing a task" (p. 616).

Recently, individuals working in the area of testing, and in language testing in particular, have begun to appreciate the roles verbal protocol analysis might play in the development and evaluation of assessment instruments (Green, 1998; Lam, 2007). Verbal protocol analysis is a complex but effective methodology. However, Crisp (2008) reiterated that Individuals choosing to use the technique require some degree of training in order to maximize the benefits in adopting this approach, and in order to avoid some of the more common misunderstandings and pitfalls associated with the use of verbal data.

In a study by Suto and Greator (2007), the verbal protocol analysis method was used to investigate the cognitive strategies utilised by examiners of General Certificate of Secondary Education (GCSE) when marking their papers. They discovered five essential marking strategies used by the examiners, namely: matching, scanning, evaluating, scrutinising, and no response. The study also revealed that the complexity of some of strategies identified confirms that GCSE examination marking

can be a difficult process, often requiring considerable expertise. Similarly, Bukta (2007) employed think-aloud protocols to uncover the decision making of five EFL teachers when marking their students' writing assignments. The study showed how such method can be used to form a better understanding of how the teachers decided on the scores of each essay.

Lumley (2002) investigated expert raters' reasoning process in assessing written works of ESL learners. Though only four raters were involved in the study, it showed that the raters followed a similar rating process. He concluded that proper training is essential in order to allow raters to learn how to interpret the task requirements and scale features so that a more accurate rating or measurement of students' performance can be achieved.

Cumming (1990) conducted a study investigating the decision-making behaviours which raters perform mentally while evaluating ESL compositions through the use of verbal protocol analysis. Based on the study's findings, the expert raters' decision-making behaviours differ from those of novice raters in terms of four broad categories: self-control focus, content focus, language focus and organisation focus. However, Cumming (1990) mentioned that the outcome of his study requires further study before a full model of the thinking process could be developed. Though Cumming's study adopted almost the same approach as the present study, it should be made clear that this study aims to investigate deeper by producing a mental model for future analysis of the raters' cognitive process. Moreover, the categories mentioned in Cumming's study are rather limited and can be expanded by providing an in-depth exploration on the raters' line of reasoning and type of knowledge states.

#### **2.4.4 Metacognitive Strategies and Skills**

King (1995) stated that “metacognition is the awareness, monitoring, and control of one’s cognitive processes” (p.16). It is the internal processing that makes use of cognitive strategies to monitor and control other learning and memory process (Flawell, 1979). Flawell (1979) maintains that metacognition consists of *metacognitive knowledge* and *metacognitive experiences or regulation*. Fox and Riconscente (2008), on the other hand, defines metacognition as the knowledge of one’s own thoughts and thought processes, which involves both conscious awareness and the capability of communicating one’s rationale. These metacognitive strategies, according to Schunk and Zimmerman (2008), allow a person to activate, guide, enhance and sustain learning overtime.

Metacognitive knowledge refers to acquired knowledge about cognitive processes, which is a knowledge that can be used to control cognitive processes King (1995). On the other hand, the metacognitive experience of a person refers to the general knowledge about how human being learns and processes information. It also includes an individual knowledge of one’s own learning processes. Based on literature, the researcher concludes that this metacognitive experience includes a person’s abilities and limitations in carrying out the cognitive activity and also his or her ability to monitor or assess the activity he is engaging in.

In relation to this, Kuhar (1998) has mentioned two components: “identifying and challenging assumptions” (p. 80). Thus, in a problem-solving task, we might add examples like weighing and assessing our judgements, choosing among heuristics or methods of problem-solving, judging whether one’s unaided skills are sufficient to the task, whether more research or a new approach is necessary. In short, metacognitive

skill 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.

In applying metacognitive theory to reading strategies, Baker and Brown (1984) have identified two clusters of activities: *knowledge about cognitive* and *regulation of cognition*. The first cluster of activities is concerned with the person's knowledge about his or her cognitive resources and the compatibility between the person as a learner and the learning situation. It is an ability to reflect on one's own cognitive processes and to be aware of one's own activities while doing the problem-solving task.

The second cluster of activities is consisted of the self-regulatory mechanisms used by an active learner during an ongoing attempt to solve problems. These mechanisms involve checking the outcome of any attempt to solve the problem, planning one's next move, monitoring the effectiveness of any attempted action and testing, revising and evaluating one's strategies for learning. Nevertheless, there is a variation of strategies used by the learner depending on the goal of the activity. For instance, reading for meaning demands different skills than reading for remembering or studying.

Thus, it can be concluded that *metacognitive* skills are skills in monitoring and controlling one's own mental processes and knowledge state, taking the cognitive skills themselves as their object (Baker & Brown, 1984). In confronting problems to be solved, writing assessors or raters are able to select and regulate the use of relevant

intellectual skills and execute task-oriented cognitive strategies. These higher order skills are termed as *metacognitive strategies*.

In the learning process, Wenden (1998) postulates that “metacognitive knowledge includes all facts learners acquire about their own cognitive processes as they are applied and used to gain knowledge and acquire skills in varied situation” (p. 82). To conclude, metacognitive strategies are skills used for planning, monitoring, evaluating the learning activity. Some of the strategies are:

- Directed attention
- Selective attention
- Self-monitoring
- Self-evaluation
- Self-reinforcement

Directed attention is employed by a learner when deciding in advance to concentrate on general aspects of a task. After that, one will use selective attention to focus on the specific aspects of a task. Then, self-monitoring is used to check one’s performance as one speaks. While self-evaluation is used to appraise one’s performance in relation to one’s own standard, self-reinforcement is used to reward oneself for success.

According to Wenden (1998), at the planning stage the learners identify their objectives and determine how they will achieve them. Nevertheless, the planning may also go on while a task is being performed. At this point, learners may change their objectives and reconsider the ways in which they will go about achieving them. At the monitoring stage, language learners act as ‘participant observers or overseers of their language learning’ by asking themselves questions like “How am I doing? Am I having difficulties with this task?” and so on. Finally when evaluating the learning

activity, there are three steps involved: first, learners will examine the outcome of their attempt to use a certain strategy in learning; second, they will access the criteria they will use to judge it; and finally they apply it.

#### **2.4.5 Problem Solving**

Problem-solving is a high level mental activity (Hunt, 1994). The first step in problem-solving is to determine the problem space and the following step is then to determine a strategy to solve it. To solve a problem, one has to reason it out and by doing this; the solver is actually progressively expanding his knowledge of the problem situation and continue doing so until he discovers the solution.

At each step in reasoning, a problem solver has to develop a mental construction of the logical situation, and then react to the features of that mental construction (Hunt, 1994). Johnson-Laird (1983) recalls this mental construction as **a mental model**. The complexity of a mental model is determined by the capacities of the immediate memory.

During the problem solving process, the solver is actually comparing his or her current state to the goal state and at the same time trying to find way(s) to eliminate the differences between the two states. If the problem shows certain familiarity, then the solver will rely to a great extent on previously memorised solution schemata. This is a short-cut through the problem space to lead to solution. Hunt (1994) states that the properly applied schemata transfer the information-processing burden from immediate memory, where the human solver is weak, to the long-term memory, where the problem solver is strong.

#### **2.4.6 The Mental Model**

Qin and Simon (1995) refer to the mental models as the structure of the subjects' knowledge about the world. They are of an opinion that a mental model is more stable, deeper, systematic, and general than the images formed for a specific task in short-term memory. It provides a source of information. Newell and Simon (1972) refer to a mental model as a problem representation. A mental model is the implementation of the different knowledge bases that enable a person (an operator) to actively gather information, make inferences, anticipate outcomes and make plans for future decision-making. It is the internal construction of the external world of some sorts that can be manipulated to make predictions and inferences.

Thus, it can be concluded that a mental model is a functional abstraction about a task/job which provides a deductive framework for problem – solving (Ryder & Redding, 1993). It contains and integrates conceptual knowledge, procedural knowledge, decision-making skills for reasoning, and strategies for problem solving. A mental model has become an important tool among training analysts and also in the system interface community. Its usefulness can be extended to school to enhance the teaching and learning process. As educators, we have to accept the fact that a person may possess several mental models regarding a single concept or problem and each mental model will represent a different view.

In formulating a mental model, a task analysis technique such as the CTA is used to collect data that yield information about the thought processes and structures underlying observable task performance (Chipman, et. al., 2000). Under this technique, VPA can be used to analyse the thinking-aloud process while a person is doing a problem-solving task, with particular emphasis on the use of representations

(the knowledge states and conceptual operations) as they justify their strategies and solutions (line of reasoning).

#### **2.4.7 Procedures in Analyzing Verbal Protocol Analysis**

The verbal protocol analysis instructions generally involved telling a person to think-aloud when he/she is performing task. Instructions may stress on talking continuously, being complete, or not worrying about saying out irrelevant things. In the process of collecting the data, the participant will be given the “keep talking” reminders after long pauses from the participant.

Verbal protocols are usually prepared for analysis and interpretation by undertaking a sequence of steps. Hassebrock and Prietula (1992), has conducted a verbal protocol analysis on a group of physician diagnosing cases of congenital heart disease. For each case, subjects read aloud the data statements from an index card and then “thought aloud” as they considered any thoughts or information relevant to a diagnosis for the patient’s condition. These verbal protocols were audio-recorded. The audio-recording of a case is converted into a typewritten transcript for the subsequent procedures of verbal protocol analysis used to identify the content and organization of medical knowledge used in the physician’s diagnostic reasoning.

For each case, the transcript included each of the groups of patient data statements followed by a new paragraph for the physician’s statements while reading the data card aloud and another new paragraph containing the subsequent thinking-aloud statements. The physicians were instructed to read the data statements on a card before thinking aloud. However they would occasionally interject a thought before

reading the complete data group. When this happened, thinking-aloud statements were included in [brackets] within the paragraph.

Syntactic breaks in thinking aloud were designated by commas, periods, or semicolons as appropriate. Pauses of short duration were designated by an ellipsis (...). For pauses longer than five seconds, they were designated by typing the word “pause” within parentheses. Abrupt changes in diction or corrections in speech were indicated by a double dash (--). Table 2.4 below shows an example of the transcription and analysis of the physician’s protocol produced in reading and thinking aloud for a single group of patient’s data.

---

21. <i>Ah—ascending aorta is slightly enlarged.</i> <sup>a</sup>
(1) Ah, well that means <u>post-stenotic dilation</u> or <u>increased ah, blood flow through it.</u>
(2) It would be consistent with ah, ah, <u>transportation of the great vessel with sub-aortic stenosis,</u> (3) uh, <u>tetralogy</u> of fallot would be ah, or a <u>tetrad variant.</u>
21.1 Hypothesis-generation: trigger $\Rightarrow$ ( <i>post-stenotic dilatation; increased blood flow through aorta</i> ) <sup>b</sup>
21.2 Hypothesis-generation: association $\Rightarrow$ ( <i>TGV with sub AS</i> )
21.3 Hypothesis-evaluation: confirmation $\Rightarrow$ ( <i>TF; Tetrad variant</i> )

---

<sup>a</sup> The first numbered paragraph typed in italics represents the physician reading the card aloud. The verbalized thinking-aloud responses follow in a new paragraph with the separate segments numbered and the relevant knowledge states identified. The coding interpretation of the thinking-aloud responses follows with the cue-number (to the left of the decimal point) and segment-number (to the right of the decimal point) noted.

<sup>b</sup> For each segment, and to the left of the arrow, there is listed the basic conceptual operation and the more specific operators used, ranging from the most general (leftmost) to the most specific operators form. The knowledge state or states in each segment are shown in parentheses to the right of the arrow, sometimes augmented with positive and/or negative signs (in the parentheses). A negative sign preceding a knowledge state shows that the physician judges the cue as being disconfirmatory evidence for the knowledge state. If there is not a symbol preceding the knowledge state, then the cue was judged as confirmatory evidence for the knowledge state. A positive and negative sign (e.g. +/-) preceding a knowledge state shows that the physician considered the cue to be both confirmatory and disconfirmatory evidence, or the cue constituted ambiguous evidence. **HH**, **PH**, **XH**, and **EH** stand for diagnostic hunches summarized after the major sections of patient data (history, physical, X-ray, and ECG) and **DX** represents the final diagnostic conclusion.

Table 2.3 Example transcript fragment and coding (Hassebrock & Preitula, 1992: p. 632)

After the audio-recording of a case is already transcribed, it was read while listening to the audio-recording in order to correct misspelled and omitted words and to correct syntactical markings. Occasionally, the physician was asked to clarify the spelling or meaning of certain medical terminology. Then a second draft of the corrected manuscript was typed and became the official transcript used in the protocol analysis, which was carried out in four sequential steps.

The following is the brief description of the steps adopted for the protocol analysis (Hassebrock & Prietula, 1992). First, the knowledge states is identified by a scorer and represented directly on the transcript by underlining them. Second, each knowledge state identified in the first step is associated with one of the possible basic conceptual operations that represent a discrete problem solving segment. Within each data group, each segment is numbered sequentially on the transcript. Third, the scorer then identifies the specific operators used to produce the knowledge state or states within the segment, based upon a judgement of correspondence between the protocol content and the operational definitions.

A physician may employ several conceptual operations for a given group of patient data. The basic operations, specific operators, and the associated knowledge states in a given protocol segment are listed in a separate paragraph below the thinking aloud transcription (see Table 2.3). Finally, the lines of reasoning (LoR) to discriminate expertise as well as individual differences between the expert physicians are identified after the analysis of knowledge states and conceptual operations.

The coding for a LoR consists of two steps: first, specific canonical models of the disease (diseases) of interest are conducted based on domain-theoretical principles.

Although most of the key canonical disease models are constructed prior to the study, unanticipated and recurring reasoning events arising during the study may necessitate the construction of other canonical model. For instance, novices may ‘be seduced’ continually by specific data to reason down a wrong path. This canonical model serves as a domain-theoretical guideline or template, to help determine specific reasoning contexts evidenced by the protocol but it does not show “how to diagnose” or “how diagnosis occurred”.

The second step of LoR coding involves reviewing the protocol in the context of the relevant canonical model(s) to determine what elements of the canonical model, and in what sequence, were instantiated by the physician’s reasoning process as evidenced by knowledge states and operators. Although the canonical model’s conceptual objects will generally be instantiated as knowledge states, differences in case data and physician knowledge can vary the sequence the elements are instantiated.

The coding for LoR reflects a summary trace of reasoning steps over a period of time in the context of a canonical model (or perhaps several models). The scorer then construct a schematic diagram (mental model) showing the data cues at which a physician evoked knowledge states found in LoR. The application of this coding scheme in studying diagnostic reasoning has revealed expert-novice differences in the use of knowledge states, conceptual operations and lines of reasoning.

## **2.5 Chapter Summary**

Section 2.1 of this chapter gave the chapter overview. Section 2.2 discussed the components in the theoretically framework of the study. The discussion begins with literature reviews on the different types of assessment in writing in sub-section 2.2.1,

and this is followed by an overview of assessment in second language writing in sub-section 2.2.2. Sub-section 2.2.3 presented a literature review on teacher feedback in second language writing, and this is followed by an overview of teacher feedback in second language writing in sub-section 2.2.4.

Sub-section 2.2.5 presented the discussion on evaluating writing, while sub-section 2.2.6, presented the theoretical framework for this study as illustrated in Figure 2.1. Section 2.3 of this chapter reviewed the related literature pertaining to the variables in this study. Sub-section 2.3.1 and 2.3.2 presented the literature reviews on the knowledge states and conceptual operators respectively as they are extensively used in the data analysis. Section 2.4 of this chapter gave a review of literature on methodology for the study. The focus of discussion is on cognitive task analysis (CTA) in writing (sub-section 2.4.1), CTA methods (sub-section 2.4.2), and verbal protocol analysis (VPA) (sub-section 2.4.3) which will be adapted for the purpose of this study.

Metacognitive strategies and skills (sub-section 2.4.4) are also described briefly as they are crucial in understanding the ‘thinking about thinking’ activity. These are related to the ‘think aloud’ activity employed in the verbal protocol analysis. Apart from that, problem solving (sub-section 2.4.5), the mental model (sub-section 2.4.6) and procedures in analyzing verbal protocol analysis (sub-section 2.4.7) are presented in detail as they are extensively used in the data analysis. This leads to the choice of research design, data collection methods and data analysis procedures for this study, and will be discussed in Chapter 3.