The paper examines some self-assessment techniques used with handicapped students and discusses the advantages and disadvantages of these techniques. The use of self-rating scales is reviewed, and questionable results are cited. Another method, in which students view an item and estimate whether they can perform it before attempting it (pre-judging), is described. It is noted that with pre-judging, tasks should be quite familiar, short, and discrete entities. Inferred self-assessment, in which students indicate they cannot perform difficult or impossible tasks, is also considered. Specific self-assessment techniques are reviewed, as are self-monitoring approaches. Among conclusions noted is that learning disabled students appear to have specific deficits in the area of self-assessment itself; the problem may manifest itself in reading, writing and spelling, math, language communication, social status, and attention-to-task skills. Studies have suggested that increased use of self-assessment may be the first step in increasing achievement. Examples of student self-assessment are appended. (CL)
FORMATS FOR ASSESSING STUDENTS' SELF-ASSESSMENT ABILITIES

Maurice Miller, Ph.D.
Tamrah Turner, B.S.

Department of Special Education
Indiana State University
Terre Haute, Indiana 47809

Presented to Council for Exceptional Children, Chicago,
April 24, 1987
What Is Self-Assessment?

Self-assessment is used to describe the act of judging one's performance on a task, with comparison to outside, objective judgements of one's actual performance. In the present context, the term is delimited to academic and academically-related skills. Further, self-assessment can be divided into: ability to gauge at all, ability to gauge in specific tasks, and ability to make judgements about one's performance. That is, self-assessment implies a stepping out of one's self and observing one's own performance in action.

Self-assessment is a component of metacognition, which involves learners' awareness of their own knowledge levels, how those levels are achieved, or how they eventually may be altered (O'Leary, 1980). Brown and Palinscar (1982) assert that metacognition has two broad elements: knowledge and awareness of one's cognition, and regulation of cognition (executive control). Currently, most of the literature relating to metacognition focuses on the second, executive control function. They note, though, that when that is the focus, the researcher or teacher must be the one to decide: (1) what the problem is, (2) how the student is doing it now, and (3), what strategy the student needs. That is, it is the responsibility of someone other than the student to
make decisions about everything leading up to the step of actually using a learning strategy.

The benefits of shifting some of this responsibility to the learners and having them develop awareness of their own cognition and performance are unproven. It seems likely, though, that increasing the student's participation would have corresponding benefits in learning (O'Leary, 1980). In fact, Loper (1984) states that the ability to predict and evaluate one's own performance is a critical metacognitive skill, particularly for children with academic deficiency.

Mildly handicapped students, particularly the learning disabled, have been described as "inactive" (Torgeson, 1982), "passive," and as evidencing "learned helplessness" (Gavalek and Raphael, 1982). Implicit in such an ascription is notice of the students' inactive stance in assessing their ability to perform and also the presumption that if the students were active in this they would be better learners.

It may be that a self-assessment continuum exists and that handicapped learners can be characterized by where they fall along that dimension. The learning disabled, for instance, by definition have sufficient cognitive capacity (i.e., I.Q.) to make judgements about their actions, but they do not do so (Torgeson, 1982). Or, they may make judgements in areas of competence but not in areas of
deficit. Another group--those we call the "mildly" emotionally handicapped--can have the capability of guaging performance but choose not to do so, possibly because of affective concerns. The amount of cognitive capacity needed to exercise metacognitive judgements--and, thus, the participation ability of the mentally handicapped--is unknown.

Assessing Self-Assessment

Identifying ways to assess students' self-assessment skills turns out to be a challenge. One cannot simply go to a student and ask, "How do you think you are doing?" Given the questionable reliability of witness reports generally (Brown and Palinscar, 1982), students' reports about their performance in anxiety-producing situations may be even less trustworthy. Also, the fact that students have had a history of failure in learning situations must cloud their metacognitions about their learning (Brown and Palinscar, 1982). Additionally, if a student's disability is in self-assessment itself, then of course one cannot ask that student to guage performance; as one child said, "I don't know what I don't know." The techniques which one must use to guage students' self-assessment abilities, then, must be either indirect or inferential.

The value of accurate self-assessment is thought to be that if the student is more self-aware, there will be a
corresponding benefit in learning. This would be the third in a three-step process:

1.) Able to accurately self-assess while engaged in a learning task,
2.) Active use of self-assessment while performing that task, and
3.) Use of self-assessment to alter one's performance.

Steps "2." and "3." cannot be investigated, though, until step "1." is developed (i.e., capability in assessing one's own performance). O'Leary says that this is an "acquired" skill and that little progress will be made until the child understands the problem that exists. Additionally, Markman (1977) notes that individuals who are willing to tolerate gaps in knowledge and inaccuracy in performance would not be expected to try to clarify this on their own. Thus, awareness is basic prior to initiating remedial procedures.

A variety of techniques have been developed to assess either the student's view of learning abilities or the student's use of self-assessment at all. The purposes of this paper are, first, to present and critique some self-assessment techniques which have been tried. Then, specific examples will be given showing how self-assessment has been used and their correspondence to assessment of actual performance. Table 1. presents a number of techniques which have been tried.
<table>
<thead>
<tr>
<th>Skill Area</th>
<th>Method</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rating self on instrument which is intended to be rated by others</td>
<td>Miller &amp; Bommarito (1983), Nathan, et al (1980), Weller &amp; Stawser (1983)</td>
</tr>
<tr>
<td>Arithmetic</td>
<td>Give 10 seconds to look over ten problems—how many can you solve?</td>
<td>Slife, Weiss, &amp; Bell (1985)</td>
</tr>
<tr>
<td></td>
<td>After computing, which ones are correct?</td>
<td>Slife, Weiss, &amp; Bell (1985)</td>
</tr>
<tr>
<td></td>
<td>Show examples of pairs of problems about two seconds (not long enough to compute) -- can you solve?</td>
<td>Schunk (1985)</td>
</tr>
<tr>
<td>Communication</td>
<td>Simply ask (e.g. Do you know what your problem is before you ask for help?)</td>
<td>Schultz &amp; Flaton (1985)</td>
</tr>
<tr>
<td>Reading Comprehension</td>
<td>Obvious incongruities (e.g. A boy makes a tuna sandwich, by mistake includes egg shells, eats it, then claims that it was good).</td>
<td>Masson (1982)</td>
</tr>
<tr>
<td></td>
<td>Imbedding inconsistencies in reading passage (e.g. non sequitur sentences)</td>
<td>Bos &amp; Filip (1982)</td>
</tr>
<tr>
<td></td>
<td>After answering comprehension questions, rate confidence in answers</td>
<td>Baker &amp; Brown (1980)</td>
</tr>
<tr>
<td>Reading Comp. OR Listening Comp.</td>
<td>Doubly self-imbedded sentences (e.g. The man, the boy, the girl knew, likes, died.)</td>
<td>Markman (1979)</td>
</tr>
<tr>
<td></td>
<td>Given directions for a game, Markman or magic trick, with a critical component missing--does child ask questions?</td>
<td></td>
</tr>
</tbody>
</table>
Have child tape oral reading and replay, listening to self, while marking errors on script (1980)

Given nonsense words:
1) Do they think they can pronounce?
2) Pronounce, then rate the accuracy of their own pronunciation attempt (1984)

Play tone every seconds and ask to check if paying attention
-then just ask at tone (not record)
-then, just have students ask selves whenever they think of it (1981)

On sociometric, have peers rate as "friend," "all right," or "wouldn't like." Then ask child how s/he thought that student rated her/him (1978)

Self-recording while studying (see "Self-Control") (1980)
Rating Scales.

In some instances, students have been asked directly to evaluate their own performance. Levine, Clark, and Ferb (1981) have developed a Self-Administered Student Profile to self-rate performance in nine areas related to school functioning. Students, and teachers and parents, had most agreement that an area was of concern in areas of Memory and Attention. There was not a great deal of agreement in the other seven areas. Additionally, items were rated along a 3-point scale and there was no objective criterion given for what each rating point meant. Reliability of the instrument has not been demonstrated nor has there been independent verification of the student's needs in rated areas. The authors point out that the instrument has diagnostic value in being a beginning-place for dialogue with the student about school performance. While that may be true, the accuracy and validity of the procedure are open to question.

That self-ratings may be questionable has been shown in two other studies. Bradley (1983) had handicapped and non-handicapped students rate themselves, and compared this to the teacher's rating, on self-image, memory, listening, directions, attention, study skills and work habits, organization, and academic achievement. While there were points of agreement, these were fewer than points of disagreement. There was more agreement between teachers and students who were not handicapped than between teachers and students who were. Bradley concluded that IQ was not,
necessarily, a good predictor of teacher-student agreement, but that affecting variables tended to be age (with corresponding level of cognitive development), cognitive style, and sex (to some extent—teacher-boy agreement higher than teacher-girl). There was not sufficient correspondence, however, to justify using self-report in anything but a general fashion. In reviewing other studies, Bradley criticized others' attempts to over-generalize from their conclusions as well as the lack of background data on instruments themselves.

Miller and Bommarito (1983) used a teacher rating scale which did have complete background data in areas of auditory, spoken language, orientation, behavior, and motor. Questions were then re-written for students to rate themselves. L.D. students were found to be close to teachers' ratings in only two areas (spoken language and behavior), and mildly mentally handicapped in three (spoken language, behavior, orientation). L.D. students generally over-estimated their skill in spoken language, orientation, and on the total score. In that study, it was interesting that L.D. and Mi.H.H. students rated themselves at comparable levels in auditory, behavior, and motor areas. Conclusions in that study were similar to those from Bradley's related to the need for reliability and validity data on the instrument in this usage. The need to clearly specify a referent group was also pointed out. The student could be rated as "average," below, or above, but it was
found that both teachers and students were uncertain as to what constituted average performance in the referent group.

In a study with mentally retarded adults using the AAMD adaptive behavior scales, Nathan, Millham, Chilcutt, and Atkinson (1980) found that these individuals could not only be accurate compared to observed performance, they were more accurate than other raters. We have found only one similar study with L.D. children (Weller and Strawser, 1983), and that did not include a complete enough report of data to be able to allow analysis. Evidently, though, there were some areas of the scale which had more agreement than others. Also, different from the Nathan, et al study, there was not independent observation to verify ratings.

Only a few studies have used procedures with handicapped children which required them to rate themselves on a rating scale and then compared that to another's rating. Those that have have found some areas of general agreement but little which is specific. Further, these studies can be criticized for lack of reliability and validity data on the instrument, over-generalizing from data, lack of independent verification of ratings, and not clarifying comparison group to the raters. It may be that this use of rating scales can be useful as a starting point in talking with the students about their own self-assessment, but current data does not support using such a scale as the self-assessment itself except in a very general way.
Pre-Judging

Another method used in studies both with arithmetic and reading recognition is having the student view an item and estimate or predict whether s/he can perform it even before attempting it. Slife, Weiss, and Bell (1985), in a carefully designed study, matched elementary-level L.D. students with non-handicapped students who were performing at the same level in arithmetic (but were chronologically younger). Subjects were given ten arithmetic problems and asked to predict how many they thought they could do. Then they worked the problems and were asked to tell which they thought were right. L.D. students were less accurate than controls in both predictions, with more likelihood of overestimating their performance. The authors use this as a basis for a tentative conclusion that metacognition really is a separate act from cognition.

A variation on this idea, reported by Schunk (1985), had L.D. students view problems but too briefly to actually solve them. Then, on a rating scale, they rated their confidence in doing these subtraction problems. While use of this procedure appeared to increase both the students' judgement ability and their actual achievement scores the study was not so clearly designed as to allow the conclusion that performance increased because of practice in predicting. Too, one would wonder why the use of the rating scale in addition simply to estimation based on the problems presented. While this looks like a usable way to get
self-assessment information in arithmetic computation, controls were missing which would allow confidence in using it.

In a study of sight word recognition, Loper (1984) used three types of non-words: high pronounceable (e.g. belmor), low pronounceable (e.g. kugafp), and nonpronounceable (e.g. dtscfk). Students were then asked if they thought they could pronounce the words. L.D. students differed from controls in judging fewer "pronounceable" words, but they did not differ on the other two types. In another study, using ranges of nonsense words, students were asked to predict, then pronounce, then judge their accuracy. L.D. students were no different from others in the judgements made, but words for them were at a lower level. Two conclusions can be drawn from this study. One relates to the way L.D. students made judgements in this kind of task—more frequently underestimating performance than overestimating it (as in the Slife arithmetic study). Also, when given tasks at ability level, L.D. students could perform comparably to non-L.D.

Pre-judgement appears to be a satisfactory way of getting self-assessment information from both learning disabled and non-handicapped students. In arithmetic, Slife, Weiss, and Bell (1985) found this to be a useful way to get data on students, and L.D. students were not as effective as other students who were on a similar achievement level, even though the comparison group actually
was younger. Schunk's report (1985) would support the conclusion that practice in pre-judging not only increases accuracy but also results in increased achievement. For reading, using a nonsense-word recognition task, Loper (1984) found L.D. students to underestimate their capabilities. However, when given tasks at their tested level, they could perform with accuracy equivalent to that of their non-handicapped peers. These studies lead to the conclusion that some kinds of tasks are more usable for self-assessment than others. Tasks appear to need to be quite familiar and also those that are short, able to be presented as single items, and scorable as absolutely "right" or "wrong." There is also a strong indication that practice increases both accuracy and actual achievement.

Inferred Self-Assessment

Several studies have had students undertake difficult or impossible tasks; when the student floundered or indicated that s/he could not do it, then the researcher inferred that self-assessment had taken place. This kind of procedure has been used with reading comprehension and listening comprehension tasks. Bos and Filip (1982), provide an example of this in asking L.D. students to find inconsistencies in print. L.D. students not only tended not to find the inconsistency, they did not indicate that an inconsistency existed. The writers' interpretation of these results was that L.D. students had a "production
deficiency"—they could have the appropriate strategies somewhere in their repertoire but failed to employ them spontaneously.

In a subsequent study (Bos and Filip, 1984) 7th grade L.D. and non-handicapped students were told to read passages which contained an inconsistency and then were asked questions, or "probes," until they identified the inconsistency. Again, L.D. students performed more poorly than did other students until the L.D. students were cued that there was an inconsistency to look for. Then, L.D. students were equally capable of finding the inconsistency. The researchers also noted that three L.D. students still failed to monitor, even with the cueing, and might be considered to have a deficit in doing self assessment at all.

In Markman's studies (1977, 1979), students were given instructions to follow or direction for doing a magic trick, but a critical ingredient was omitted. Students were scored as responding appropriately when they indicated that something was missing or that they did not have complete information. Markman found that reading disabled students simply did not recognize the omission automatically. In fact, they would continue attempting the magic trick, for instance, even when it was quite evident that it could not work. However, when students were cued to look for the omission, they could find it. Markman points out that such students evidently are willing to tolerate knowledge gaps
and not seek clarification, and awareness of one's own comprehension failure provides a basis for remediation activities.

These studies lead to several conclusions. In more complex, verbal tasks, students with learning handicaps are quite willing to tolerate confusion, ambiguity, and outright error. In one report (Bos and Filip, 1982), when given passages which did not make sense, one L.D. student rated it as "good" because, he explained, "The stories I usually read don't make sense" (p, 82). These students do not automatically use, or even seek to use monitoring tactics. At the same time, there is evidence that when these students are told to look for a particular kind of error, they can do so. They just do not do it spontaneously, even when it is obviously needed. There is also a strong inference by most of these researchers that increased self-assessment would result in increased achievement.

**Specific Self-Assessment**

Varying results have been obtained when students have been asked to do task-specific self-assessment activities. Miller (1986) described use of a picture of a stoplight and the student being asked to put a finger on a light when doing each item; green meant "I think I'm right," red meant "I really think I'm responding incorrectly," and yellow meant "I really don't know if I'm right or not." Given such tasks and visual and auditory perceptual tasks, word
recognition, and sound blending, L.D. students could do the self-assessment task, but their accuracy increasingly faltered. Some students erred by over-estimating their performance (finger on green for an incorrect item) and some by under-estimating (finger on red but getting the item correct, in fact). Both patterns were shown and a particularly L.D. pattern did not appear. When intermediate-age L.D. students were compared to intermediate-age mildly mentally handicapped students, the Mi.M.H. students were significantly better at self-assessment, even though their actual performance was at a lower level.

Bruininks (1978) has reported similar results in the area of social status. Classes in which L.D. students were mainstreamed were asked to do sociometric ratings for each student on the class roster, and L.D. students were then asked to assess their social status by stating how they thought each student rated them. L.D. students were inaccurate—thinking that their classmates had rated them much higher than they actually had. Bruininks also states that, when other students were asked to do the same thing, they tended to underestimate their actual rating. Garret and Crump (1980), with a similar finding, interpret this as modesty which, then, is also a social grace the L.D. students did not have.

Harris and Graham (1985) point out that self-assessment and awareness by themselves may not have an impact on the
student's learning and achievement. However, when this is coupled with specific strategies an achievement effect can be noticed. They demonstrated that learning disabled students can be taught to follow a model to score their own written compositions and graph results. This was then coupled with specific learning strategies with a resulting increase in written composition performance.

"Self-monitoring" appears to be a subset of specific self-assessment, and this has been used in several contexts. The usual format is to have a bell or notification at fairly regular intervals (e.g., averaging every 45 seconds), and students are to check whether they are on-task or off-task.

Stoller, (1980), for example, studied self-monitoring effects on spelling when L.D. students simply monitored and recorded whether or not they were studying. Even though nothing more task-specific than that was emphasized, there were positive increases in number of spelling words learned, at least on a short-term basis.

More typically, self-monitoring has been used for attention-to-task in a variety of contexts. Hallahan, Marshall, & Lloyd (1981) note that self-monitoring requires both self-assessment and self-recording. They demonstrated that students could record and that this self-monitoring did increase attention. Additionally, they showed equal success when they played a tone, played a tone but did not require
recording, or reminded students just to ask themselves if they were paying attention "whenever you think of it."

Heins (1980), on the other hand, found considerably less attention to math if students were not cued.

Hallahan and his colleagues have investigated a number of specific variables related just to the way self-monitoring and self-recording work most effectively. They found (Hallahan, Marshall, and Lloyd, 1981) that accuracy in self-recording was not as important as the act of recording itself, though "a certain level of ... accuracy is necessary" (p. 412). On the issue of whether or not recording was actually necessary (Lloyd, Hallahan, Kosiewicz, and Kneedler, 1982), results were mixed. They concluded that the need for recording probably varies from setting to setting. They also showed that self-recording techniques could be implemented in the mainstream classroom with good results—particularly when the student was reinforced for recording (Rooney, Hallahan, and Lloyd, 1984).

Results on the use of specific self-assessment are mixed, probably due both the variety of contexts in which it has been used as well as the variety of ages and types. Miller (1986) found that increases in task difficulty led to decreases in self-assessment accuracy in L.D. students—in fact, not performing as well as the mildly mentally handicapped. Studies of self-assessment of social statues (Bruininks, 1978; Garret and Crump, 1980) have also found
the L.D. student to be less accurate than mainstream peers. There is some evidence that self-assessment by itself does not lead to productivity and needs to be coupled with self-recording, as in the Virginia Institute studies, as with a learning strategy for remediation (Harris and Graham, 1985).

**Strategies**

Other procedures to use for self-assessment have been suggested, though they lack a research basis. Pflam and Pascarella (1980) mention a reading method of having a child tape record a passage read orally, then listen to the tape and mark the "script" copy for errors. In some areas, self-assessment may be done simply by asking the students how they think they are doing. Schultze and Flaton (1985) have used this in the area of oral communication. Safran and Safran (1982) and Weller and Strawser (1983) both appear to think that there is usefulness in students filling out self-rating scales. Their point seems to be less in terms of accuracy, however, than that this presents a starting point for communication with the student since the student has just indicated self-perception.

There may also be a number of strategies which can be taught to students to indicate their self-assessment accuracy. These include such things as self-questioning, visual imagery, and various monitoring and recording
techniques (Bos, 1983; Cooke and Slife, 1985; Leone, 1983; Lovitt, 1984).

Discussion

Self-assessment is a metacognitive act. It requires the users to observe their own performance and render some judgement about that performance. With the current emphasis on metacognition and related learning strategies in the field, it might be thought that self-assessment has already been thoroughly addressed. This, however, is not the case. O'Leary (1980) points out that success for cognitive training is improbable unless there is also training in the self-awareness and assessment areas. Several other writers note that use of a number of teaching techniques which do not involve the learner in the assessment and decision-making are, in effect, increasing the student's passivity, the student's lack of learning skills, and simply too much teacher involvement in place of the student's (Brown and Palinscar, 1982; Harris and Graham, 1985; Kneedler and Hallahan, 1981). These all point to the need for attention to self-assessment and its role in learning, particularly for students with learning handicaps.

Because learning and self-assessment of learning are covert, it is difficult for the professional teacher or researcher to know just how the student is assessing performance or if, in fact, there is any self-assessment taking place. This paper has reviewed a number of devices
and techniques which have been explored for gauging the student's self-assessment. While there have been a variety of findings, some conclusions are possible. First, learning disabled students appear to have specific deficits in the area of self-assessment itself. This deficit may be general or specific to the achievement area where the student has most difficulty. It has been shown, however, that the deficit may manifest itself in a number of areas: reading, writing and spelling, math, language communication, social status, and attention-to-task skills. Several studies have demonstrated that students do have the capability to understand the task and assess their performance in it, but they fail to employ them spontaneously. These studies have shown, though, that at least some students can rapidly increase assessment accuracy if reminded to do so.

Research has also demonstrated that, in some contexts, increases in utilization of self-assessment brings corresponding increases in achievement. In those studies which have found that self-assessment by itself does not result in achievement, it has been presented as the necessary first step. This paper has, then, presented a number of ways in which students' use of self-assessment can be assessed. Corresponding techniques for teaching self-assessment capabilities have also been presented. Several different examples of students' problem areas, use of self-assessment, and interpretations of the student's performance are presented in the Appendix.
References

General


**Assessment**


Attention, Self-Control, and Self-Monitoring


Communication

Mathematics


Reading - Comprehension


**Reading - Decoding**


**Social**


**Spelling and Writing**

Strategies


APPENDIX

Self-Assessment Examples
Student Self-Assessment Examples

Student: Kris    Age: 11    Grade: 4
Learning Characteristics:
In reading passages containing suffixes, drops suffixes, changes suffixes, or changes root words keeping correct suffixes.
Method:
Reading aloud a written passage and indicating by pointing to the word she did not know
Results:
Corrected a suffix omission error
Indicated a mispronunciation error
Indicated uncertainty about a word’s meaning
Committed a suffix error – changed the root word

Student: Kim    Age: 13    Grade: 5
Learning Characteristics:
In the area of written expression, inconsistencies in grammar and sentence structure
Method:
Unscrambling a series of words into a meaningful sentence and stating whether or not she was happy with the way the sentence sounded
Results:
Failed to recognize or acknowledge any errors

Student: Diane    Age: 11    Grade: 5
Learning Characteristics:
In writing spelling words, omits or substitutes the prefix or suffix, gets prefix or suffix correct but changes root word, or switches position of the prefix with the suffix
Method:
Stoplight
Result:
Failed to indicate any errors

Student: Harold    Age: 8    Grade: 2
Learning Characteristics:
Inconsistency in utilizing word attack skills and difficulty in recalling previously learned words
Method:
Before reading, scan the page and point to any words he thought he could not pronounce
Stoplight
Results:
Accurate in identifying words he thought he could not pronounce
Failed to identify all the words he thought he could not pronounce
Never indicated that he mispronounced a word
Indicated uncertainty about some words he mispronounced
Indicated uncertainty about some words he pronounced correct
Placed finger on yellow while sounding out a word – moved finger to green when he had sounded out the word correctly and knew it