low accuracy on this section and also may be less stable than
the high-risk group. This low level of accuracy may indicate
that primitive reflexes are less easily defined in the low-risk
group and more variable in appearance over short periods of
time.

A strong implication of the data on tone section is that
raters tend to disagree on the risk status of muscle tone
although they are relatively stable in their own estimates over
a period of time. The trends in the data suggest greater
interobserver agreement for the high-risk infants than for the
full-term infants with regard to muscle tone.

The section on volitional movement demonstrates poor
accuracy for the total-risk score, which is rather surprising.
We had expected that items that incorporate less infant han­
dling, as is true of many of the items in this section, would
be easier to score. The opportunity for a 4-month-old infant
to get a risk score on this section is much less, however, than
in the other sections, so that the overall variability of scores
is low. This fact dramatically affects the stability coefficient.
In this section, agreement is relatively high between subse­
tuent test administrations, and inspection of the data indi­
cates that the extremely low occurrence of abnormal or ques­
tionable scores of the low-risk group may create an artificially
low estimate of the stability.

CONCLUSION

The purpose of this article was to present reliability data
for the MAI in assessing 4-month-old infants. The data pre­
 sented suggest that both accuracy (interobserver reliability)
and stability (test-retest reliability) of the MAI total-risk scores
are fair when looking at a combined group of low-risk and
high-risk infants. The accuracy and stability for the individual
section-risk scores were also presented, and implications of
these data were discussed.

Because of the relatively small sample sizes, particularly in
assessing test-retest reliability for the high-risk infants, ad­
tional reliability data need to be collected. Furthermore, stud­
dies to examine the reliability of individual items on the MAI
will be conducted in an effort to improve or clarify criteria
used for those items with low reliability. In addition, predic­
tive validity studies designed to examine the success of the
MAI in the early detection of cerebral palsy will be carried
out.

The findings reported suggest that the MAI has potential
for use as both a clinical and a research instrument by
pediatric physical therapists, pediatricians, nurses, and other
health professionals who are working with high-risk infant
groups. Because the MAI is one of the few comprehensive
neuromotor infant assessment tools that has systematically
been examined for both interobserver and test-retest reliabil­
ity, this tool has the capability for becoming a valid instrument
to assess the motor abilities of young infants who are at risk
for movement disorders.

Acknowledgments. The authors are grateful to Andrew
Hummel for his support and advice in data programming and
analysis.

REFERENCES

1. Chandler LS, Andrews MS, Swanson MW: Movement Assessment
   1980, pp 3-53
2. Kerlinger FM: Foundations of Behavioral Research—Educational and Psy­
   chological Inquiry. New York, NY, Holt, Rinehart & Winston General Book,
   1964, pp 429–443
   Psychological Corporation, 1969, pp 1–178
4. Hartmann DP: Considerations in the choice of interobserver reliability
   1978, p 125
6. Currier DP: Elements of Research in Physical Therapy. Baltimore, MD,
   Williams & Wilkins, 1979, p 159
7. Atwater SW, McEwen IR, McMillan J: Assessment of the reliability of
   pediatric screening: A tool for occupational or physical therapists. Phys
   Ther 62:1265–1268, 1982
8. Ferguson GA: Statistical Analysis in Psychology and Education. New York,
9. Werner E, Bayley N: The reliability of Bayley’s revised scale of mental and

Commentary

The authors of this article are to be commended on their
attempts to provide reliability studies for the Movement As­
essment of Infants (MAI). The development of an assessment
tool by physical therapists and the necessary documentation
to allow its acceptance by other disciplines is certainly a major
undertaking. The paper’s purpose is to examine systematically
the reliability of the MAI as a tool, and I believe that the
authors have fulfilled their intent. Several questions should
be raised, however, about the extent of the test’s reliability.

The issue of reliability of an assessment tool is an important
one. General agreement among statisticians is that a test
should have a reliability coefficient of at least .94.1 Others,
however, believe that a minimum of .90 is acceptable. The
higher standards of above .90 are rarely attained, and many
useful tests have reliability coefficients in the .80s or slightly
below. The lowest reliability coefficient that is acceptable is
usually .70.1

The MAI was developed for the purpose of assessing motor
behaviors of high-risk infants. If only the reliability coeffi­
cients for accuracy in the assessment of the high-risk infants
in the study are examined, only one of the risk scores reaches
the acceptable value of .70 for reliability. In fact, if all scores
(high risk and low risk) are examined, only one risk score
reaches the .70 value. Does this study actually give enough
evidence that the MAI is a reliable tool in the assessment of
the group of infants for which it was developed? As the authors
point out, the stability coefficients or test-retest scores are
higher than the accuracy coefficients. I would expect these
scores to be higher, however, because the same person is
repeatedly giving the test, and intrarater rather than interrater
reliabilities are being assessed. The scores for the stability
items are additionally for a small number of children, and as
the authors state, a larger sample size is needed to determine
the stability of the trends noted.
The article raises several questions in my mind concerning the methodology used. A major concern is the sample on which the testing was done. The infants were predominantly white, and no blacks, hispanics, or other ethnic groups, except for one Asian, were represented. Researchers continually raise questions on the effects of race and culture on the attainment of certain developmental tasks and, in particular, on the development of gross motor skills. Although this study is not directed at documenting the validity of the tool, a clinician must be assured that the controls for establishing internal validity have been applied both in this study and in the development of the tool before accepting its validity. The question to ask is if an adequate cross section of races, sexes, socioeconomic factors, and cultures have been used in the development of the tool. Validity is usually thought to be of more value than reliability because a test can be sufficiently valid without being reliable. Being reliable does not make a test valid, however, but only contributes to the overall validity.

As stated by the authors, the MAI has not been empirically normed. The lack of these norms is a major flaw in the degree of acceptance of this tool among other disciplines. Even with the degrees of reliability found in the study, questions will continue to be raised until normative data on both full-term and preterm infant groups are collected and analyzed. According to Campbell, studies on both of these groups must be done if the appropriateness of the MAI for its stated goals is to be demonstrated.

The authors point out in the discussion of their testing procedures that the testing situations for the high-risk and low-risk infants might be a confounding variable when comparing group differences. I believe that this is an important variable to be considered. The test results of the MAI may rely heavily on the skill and experience of the test administrator in maintaining the interest and cooperation of the infant. In this study, the examiners were experienced and adequately trained in the administration of the MAI and yet reported difficulties in maintaining the infant's interest. Therapists with less experience and less training with the MAI most likely will experience more difficulty with their interpretation of the results. The administration of the test may thus yield less reliable results.

The authors are cautious in their interpretations of the reliability findings in the discussion of their results. They have been clear in their reporting of many scores that represent poor to fair intrarater reliabilities and test-retest reliabilities and of the few scores that represent good reliability. None of the reliability scores reached the high levels that are desired in the validation of assessment tools. The lack of attaining the desired levels of .80 and above seems to raise questions about the reliability of the test and thus dampens the conclusion of the authors that the MAI has attained reliability.

I applaud the authors in their discussion of the limitations of the study (the homogeneous sample, the sample size, and the intervening variables such as the simultaneous administration of the Bayley Scales of Infant Development). In recognizing these limitations, further studies on the reliability of the tool and on the validity of the tool by these authors should prove to have fewer questions raised about the validity of the test. This study should prove to be an excellent pilot investigation for the future studies planned by the authors on examining the reliability of individual items on the MAI and on the predictive validity of the instrument.

I agree with the authors that the findings suggest the possibility of the use of the MAI as a clinical and as a research instrument. I do not believe, however, that the results of this present study are strong enough to support a claim that the MAI is a reliable instrument. The low reliability coefficients, particularly in the area of accuracy, are a serious concern.

In summary, this study represents an excellent initial attempt at examining the reliability of the MAI. Further study is needed to answer the many questions that can be raised regarding not only the reliability but also the validity of the MAI. The results of these future studies should help to determine the appropriateness of the MAI in both clinical and research settings.

BARBARA H. CONNOLLY EdD, Associate Professor University of Tennessee Center for the Health Sciences Program in Physical Therapy 800 Madison Ave Memphis, TN 38163

REFERENCES