Validity of the Developmental Checklist of the Developmental Observation Checklist System

Monte Gannon
Western Kentucky University

Follow this and additional works at: http://digitalcommons.wku.edu/theses
Part of the Education Commons, and the Psychology Commons

Recommended Citation
http://digitalcommons.wku.edu/theses/397

This Thesis is brought to you for free and open access by TopSCHOLAR®. It has been accepted for inclusion in Masters Theses & Specialist Projects by an authorized administrator of TopSCHOLAR®. For more information, please contact topscholar@wku.edu.
VALIDITY OF THE DEVELOPMENTAL CHECKLIST OF THE DEVELOPMENTAL
OBSERVATION CHECKLIST SYSTEM

A Thesis Proposal
Presented to
The Faculty of the Department of Psychology
Western Kentucky University
Bowling Green, Kentucky

In Partial Fulfillment
of the Requirements for the Degree
Specialist in Education

By
Monte Hugh Gannon

August 2007
VALIDITY OF THE DEVELOPMENTAL CHECKLIST OF THE DEVELOPMENTAL OBSERVATION CHECKLIST SYSTEM

Date Recommended 6/29/07
Elizabeth Jones
Director of Thesis

Richard A. [Signature]
Director of Graduate Studies 10 July 2007
Acknowledgements

I would like to take time to thank those people who have helped and influenced me along the way. First, I would like to thank my thesis committee, Dr. Carl Myers, Dr. Vicki Stayton, and Dr. Elizabeth Jones. I would like to thank them for the time, effort, and expertise they provided for this project, especially Dr. Jones as the chair of my committee. I would also like to thank my parents and other family members for the excellent guidance, love, and support they have given me in my life. Finally, I would like to thank God, without whom nothing is possible.
Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Tables</td>
<td>iii</td>
</tr>
<tr>
<td>Abstract</td>
<td>iv</td>
</tr>
<tr>
<td>Literature Review</td>
<td>3</td>
</tr>
<tr>
<td>Methodology</td>
<td>24</td>
</tr>
<tr>
<td>Results</td>
<td>28</td>
</tr>
<tr>
<td>Discussion</td>
<td>38</td>
</tr>
<tr>
<td>References</td>
<td>44</td>
</tr>
<tr>
<td>Appendix A</td>
<td>48</td>
</tr>
<tr>
<td>Appendix B</td>
<td>50</td>
</tr>
</tbody>
</table>
## List of Tables

<table>
<thead>
<tr>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptive Statistics for the DC of the DOCS and the BDI-2</td>
<td>29</td>
</tr>
<tr>
<td>Correlations of DC of the DOCS Domains with BDI-2 Domains and Total Score</td>
<td>30</td>
</tr>
<tr>
<td>Classification Agreement for DC Domains with BDI-2 Domains 2 SD</td>
<td>32</td>
</tr>
<tr>
<td>Classification Agreement for DC Domains with BDI-2 Domains at 1½ SD</td>
<td>34</td>
</tr>
</tbody>
</table>
This study investigates the use of the Developmental Checklist (DC) of the Developmental Observation Checklist System (DOCS; Hresko, Miguel, Sherbenou, & Burton, 1994) as a screening measure for children birth to three years of age. Kentucky regulations require the use of the DC for screening in the KEIS-First Steps program, yet there is little information regarding the effectiveness of the measure as a screening tool. Evaluation data are presented for 51 children between the ages of 2 and 35 months on the DC and the Battelle Developmental Inventory, 2nd Edition (BDI-2; Newborg, 2005). The DC and BDI-2 domains evidence no significant correlations for this sample. However, significant moderate to strong intracorrelations were evidenced within each measure. Analysis of contingency table components using the BDI-2 as the criterion measure yield specificity (percentage of true negatives) and sensitivity (percentage of true positives) for cutoff standard scores of 94, 85, and 77 on the DC. The number of accurate referrals and accurate nonreferrals comprise the classification consistency between the two measures. Adequate sensitivity (above 80%) is evidenced at the most liberal cutoff for the DC (standard score ≤ 94) at both BDI-2 placement criterions (1½ and 2 SD below the mean) for all domain comparisons. However, there is a cost of less than adequate specificity at
this level (94 or below). Good sensitivity was also evidenced for one additional comparison, the Adaptive/Social comparison at the 2 SD BDI-2 criterion with the 85 cutoff on the DC. Classification consistency was not evident at 1 ½ SD on the two domains criterion on the BDI-2 for even the most liberal DC cutoff score (standard score ≤94). However, sensitivity levels are adequate for all domain comparisons at the liberal cutoff standard score of ≤94. The results indicate concern for the use for the DC with the BDI-2 using anything but the most liberal cutoff on the DC. A discussion of practical implication for use of the DC questions the use of the DC with the BDI-2 and provides suggestions for further research with the DC.
Literature Review

In our present society, early intervention for children birth to three years with developmental delays and disabilities has become increasingly important. The importance of early intervention is apparent in research results that indicate how vital it is to identify developmental problems and delays early in childhood to facilitate timely intervention. Short- and long-term benefits of early intervention are evident for individual children, their families, and society in terms of social, behavioral, and academic outcomes (RAND Corporation, 2005). Accurate identification of young children in need of early intervention is vital and the focus of the present investigation.

This literature review will first examine the need for and importance of early intervention. The review provides legislative rationale for early intervention along with state regulations and procedures. Next, the document presents procedures for identifying children birth to three years in need of early intervention services, along with a review of the Developmental Checklist (DC) of the Developmental Observation Checklist System (DOCS; Hresko, Miguel, Sherbenou, & Burton, 1994). The DC is the focus of the present research. The DC is a screening test for detecting developmental delays required by Kentucky Administration Regulations (KAR), 911 KAR 2:110 (Kentucky Cabinet for Health and Family Services, 2006a), for use in the screening process for identification of children potentially in need of early intervention services in Kentucky. This investigation examines the concurrent validity of the DC in a referred sample from one First Steps program in Kentucky. Also examined is the classification consistency, including the sensitivity and specificity, of the DC. This information is critical to an understanding of
the validity of the DC, yet no information is currently available regarding these properties (Glascoe, 1991).

Efficacy of Early Intervention

Early identification of and intervention for children with developmental delays and disabilities has been shown to produce many short-term and long-term benefits for children, families, and society. Such benefits include academic or cognitive gains, improved high school graduation rates, greater success in the labor market, behavioral benefits, a decrease in the amount of future crime, and diminished amounts of hospitalizations and injuries (RAND Corporation, 2005). Additionally, society benefits from these gains economically (Schweinhart, Barnes, Weikart, Barnett, & Epstein, 1993).

One benefit of early intervention that has emerged from several studies is a gain in academic performance or cognition. In a study at the University of Pennsylvania Children’s Hospital, the researchers compared infants and toddlers who were in need of early intervention services and either received the services or did not receive the services. The intervention group who received the early intervention services had significantly higher increases in IQ than the nonintervention group, with a mean increase of 8.1 points versus 0.8 points, respectively (Goodman, Cecil, & Barker, 1984). The RAND Corporation (2005) reviewed empirically supported outcomes of early intervention and compiled a list of early intervention programs (e.g., Abecedarian and Early Head Start), which have evidenced cognitive or academic gains for program participants. RAND researchers identified studies that document long-term outcomes such as improved high school graduation rates and increased levels of future success in the labor market. These long-term gains often have a huge impact on the children, their families, and society. The
studies reviewed in the RAND Report indicated the potential for individual growth and advancement in early childhood, if the proper circumstances are present. Another outcome documented in the research on the benefits of early intervention for young children with special needs is behavioral benefits. The behavioral benefits documented include increases in positive behavior and social competence, among others (RAND Corporation, 2005). Additionally, long-term follow-up of individuals who have received early intervention shows significantly lower crime rates in this population. Other studies have even suggested diminished amounts of hospitalizations and injuries for children who have participated in early intervention programs. Overall, the RAND report concluded that early intervention services lead to higher functioning in many areas for children (RAND Corporation, 2005).

If children with special needs function on a higher level because of early intervention, then early intervention services should actually save money in the end. Hypothetically, the money spent on early intervention is an investment for the future. Children who function on a higher level will require fewer future services, thus less money in the future, because they will have decreased needs for overall support. For example, providing an institutional placement for an individual with special needs for an entire lifetime is very expensive. However, financial savings are evident if early intervention services enable some children with special needs to grow into self-sufficient adults (Bowe, 2004). The Perry Preschool Project is an example of a program that displayed this point (Schweinhart et al., 1993). This project demonstrates that when children have appropriate services, the expectation is for them to earn more money over a lifetime, thus saving money in the end. Additionally, an impressive finding from the
Rand Report showed that early intervention programs could produce a return to society anywhere from “$1.80 to $17.07 for each dollar spent on the program” (RAND Corporation, 2005, p. 1). Therefore, the money initially spent on these programs actually becomes a financial investment for the future.

Other studies demonstrate similar positive outcomes for recipients of early intervention. A recent study evaluated children attending state-funded preschool programs in five states and compared them with children who were not attending a preschool program. The students involved in the preschool programs made significant gains in language, literacy, and mathematical development, while those children not involved made no such gains. The gains in the children who attended the preschool program were evident regardless of economic standing or ethnic background (National Institute for Early Education Research, 2005).

Early intervention is also vital neurobiologically and developmentally due to the plasticity of young brains. Plasticity is a term used to describe the ability of the brain to recover from injury. Evidence supporting the brain’s ability to recover function or adapt structures or functions subsequent to trauma or insult indicates that this is more likely to occur early in the developmental period. For example, greater functional sparing is documented in a study of individuals with hypoxic-ischemic events before the age of one versus those who had similar problems between the ages of 6 and 14. The study indicated children who have such injuries at earlier stages of life tend to have greater functional sparing than those injured later, the former group of children performing better on a test of immediate memory than the latter. The authors interpreted the results to support that the younger the children are, the greater plasticity of their brains, or ability to
overcome damage and injury (Vargha-Khadem, Salmond, Watkins, Friston, Gadian, & Mishkin, 2003). Young children are better able to recover neurobiological functioning after injury due to the greater number of neurons (Shonkoff & Marshall, 2000). A two-year-old has more synapses (neural connections) than an adult has. Establishment of neural pathways or pruning of unused neural pathways occurs as young children mature. This pruning process continues until puberty. By puberty, an adolescent would have half as many synapses as when that child was eight-months-old (500 trillion versus 1,000 trillion, respectively). Interventions will be more effective the earlier they take place, while the brain is the most adaptable due to the larger number of available neurons to establish paths or connections, making the birth to three age range a critical period in development (Chen, 1999). Thus, identification of children with special needs is most beneficial at a young age.

The research reviewed supports the benefits of early intervention for children in terms of developmental, academic, and cognitive gains, for families in terms of their children's behavioral and social gains, and for society in terms of economic gains. Since these services have such substantial and pervasive benefits, it is important to determine those in need of early intervention services. The next section will review the process for determining who receives early intervention services.

Legislation

Research data supports the efficacy and vitality of early intervention and serves as the impetus for legislation to ensure the provision of early intervention services. Public Law 94-142, the Education of All Handicapped Children Act of 1975 (EHA) mandates that children ages three to 21 years with disabilities receive a free and appropriate public
education. For children with disabilities, the legislation requires that appropriate education be implemented through the use of individualized educational plans (IEP’s) and that parents be involved in the writing of the IEP’s. Also mandated is that services be implemented in the least restrictive environment (Shonkoff & Marshall, 2000). Public Law 94-142 has been amended several times. One important revision took place in 1986 (under P.L. 94-457), when changes were made to one section (Part H) to increase service age ranges for at-risk children to include the birth to three age range if a state enacted legislation to provide services (Nuttall, Romero, & Kalesnik, 1999). Although states are not required to serve preschool children (three to five years of age) with disabilities, states received grant money if they chose to comply (Chen, 1999). In 1990, the reauthorization of EHA renamed it, the Individuals with Disabilities Education Act (IDEA). The 1997 amendments to IDEA continued the mandate that services be delivered to individuals in the birth to three age group, only now under Part C (formerly under Part H). In addition to components about the assessments and service plans that should take place, Part C also mandated that states develop child find systems to locate all children potentially eligible for early intervention services (Nuttall et al., 1999). The Individuals with Disabilities Education Improvement Act of 2004 (IDEIA) is the amended IDEA. However, it will be another year or two before the 50 states will be able to create implementation regulations from these current amendments. Therefore, the Part C regulations for the 1997 amendments are the regulations in current use.

As mentioned above, Part C of IDEA (Part H from 1986) addresses the needs of children from birth to age three. Grant funding to encourage states to serve children who are under the age of three, and in need of early intervention services, is available for
states who voluntarily comply. Services extend to children birth to three who meet the criteria for a diagnosed condition, an established risk, or a developmental delay.

Developmental delay is an important concept. When used in the early intervention field it describes children who do not have a clear diagnosis yet, but are clearly behind in developmental milestones (Bowe, 2004). To ascertain whether services are needed, a screening of the child’s developmental status is conducted to identify the need for further evaluation. If the screening results indicate the need for further evaluation, then an evaluation that measures the child’s developmental status in the five mandated developmental domains (cognitive, language, motor, social-emotional, and adaptive) is conducted. If the evaluation reveals a significant delay in one or more of the five domains, then the child is eligible to receive early intervention services. This study is primarily concerned with the assessment instrument used for this screening process.

**KEIS Screening and Evaluation Services**

As is true for each of the fifty states, Kentucky is responsible for developing and implementing state regulations to comply with Part C of IDEA. In order to comply with Part C, Kentucky has passed legislation to create an early intervention program, called Kentucky’s Early Intervention System – First Steps (KEIS – First Steps). Kentucky Administration Regulations (KAR; Title 911, Chapter 2) provide regulations for the provision of services for children aged birth to three to comply with federal directives. Further, they establish the procedure for referral of children to KEIS – First Steps to one of the regionally based offices called Point of Entry (POE) when there is a perceived developmental problem or issue. The POE staff provides initial screening and referral for reduced or no cost evaluation services. The Kentucky Administration Regulations
(KAR), 911 KAR 2:110, provides statutes for the POE staff to follow in the referral process. Section 1 (6) (c) 4 dictates that if a referral is appropriate, then an approved screening test must be administered (Kentucky Cabinet for Health and Family Services, 2006c).

Specifically relevant to this study, in January 2005 revisions to section 1 (6) (c) 4, specified that the screening instrument utilized must be the Developmental Observation Checklist System (DOCS; Hresko et al., 1994; Kentucky Cabinet for Health and Family Services, 2006a). Often, the Developmental Checklist (DC) alone is administered, instead of the entire DOCS, according to KEIS Evaluation Specialist N. Slaughter (personal communication, October 9, 2006). In addition, although scores for the four DC domains are typically calculated, the DC Total score is often not determined. Once the DC is administered, typically via interview of the child’s parent, it is then determined if the scores are indicative of risk factors that should be further evaluated and addressed. A risk factor(s) is (are) identified when one or more of the domains on the DOCS appears to be significantly low, indicated by a standard score of 94 or below on any of the identified domains of the DOCS (Kentucky Cabinet for Health and Family Services, 2006a). If the scores on the DOCS indicate a potential problem or risk, a referral for an evaluation, called the Primary Level Evaluation (PLE) ensues. If the scores from the PLE indicate there is no risk or potential problem, then the family receives the recommendation to return to the POE if the problem(s) persist after three additional months. In addition, families receive information about other resources that may address their referral concerns. However, if the parents still insist that a problem exists and do not want to
wait three months, then the evaluation process is initiated (Kentucky Cabinet for Health and Family Services, 2006b).

A recommendation for a PLE occurs if the results from the DOCS screening indicate that risk factors are present. A Primary Level Evaluator, an individual approved by the state to complete the evaluation, conducts the evaluation. A Primary Level Evaluator must have a Bachelor’s degree in a related field and often a Master’s degree as well. Additionally, experience with children birth to age two, children with disabilities, and norm-referenced assessment tools is required (Kentucky Cabinet for Health and Family Services, 2006d). A PLE involves a five-area evaluation, as per 911 KAR 2:120, Section 1 (4) (c) that includes the following developmental areas: cognitive, motor, language, social/emotional, and adaptive. The PLE does not require the use of any specific assessment instrument. However, because the assessment tool must be norm referenced and assess the five developmental domains, the Battelle Developmental Inventory, 2nd Edition (BDI-2; Newborg, 2005) is frequently used. Few other tests meet such criteria. If other norm referenced instruments are used, then more than one test may need to be given to comply with 911 KAR 2:120, Section 1 (4) (c), making it much easier to choose the BDI-2, which solely meets all of the requirements. Currently, the BDI-2 has become a typical choice for the PLE in Kentucky, according to N. Slaughter (personal communication, October 9, 2006). After the evaluation is complete, the Primary Level Evaluator is responsible for making a determination of eligibility based on the results of the testing, as per 911 KAR 2:120, Section 1 (4) (e) 2 (Kentucky Cabinet for Health and Family Services, 2006b).
To be eligible for services under the Developmental Delay classification, performance on the norm-referenced diagnostic measure must meet one of two criteria. One criterion would be a score falling at least two standard deviations below the mean in at least one developmental area. The other criterion for qualification is a score of at least one and a half standard deviations below the mean in two or more developmental areas (Kentucky Cabinet for Health and Family Services, 2006b). Since intervention services derive from assessment results, it is imperative that these instruments be psychometrically sound.

After the PLE, a meeting takes place to discuss the results of the evaluation and determine the need for services. If a significant Developmental Delay is present, then an additional meeting must happen to ensure appropriate services will be in place. Section 1 (2) (a) of KAR 2:120 indicates that an Individual Family Service Plan (IFSP) must be conducted within forty-five calendar days from the initial referral date. The initial IFSP meeting determines what strategies and activities will meet the outcomes desired by the family. Additionally, the IFSP meeting outcomes spell out what supports and services are necessary to meet the desired outcomes.

To summarize, Kentucky’s regulations and procedures require the use of the DOCS as the screening instrument at the POE. A Primary Level Evaluator conducts a comprehensive assessment if the score on the DC of the DOCS indicates a potential problem or risk factor (standard score of 94 or below). The BDI-2 is a preferred instrument for use as the comprehensive assessment tool because it assesses all of the mandated domains while being norm referenced. Test results from the comprehensive assessment tool and the recommendations by the Primary Level Evaluator determine
whether children are eligible for services. Kentucky mandates the use of the DOCS; therefore, it is important to examine the instrument’s ability to appropriately screen and identify children in need of further evaluation.

Developmental Observation Checklist System (DOCS)

The Developmental Checklist of the DOCS (Hresko, et al., 1994) is required for use as the screening tool by the POE staff to determine whether a child should receive a PLE. This section will review the data available regarding the psychometric adequacy of this instrument. Alfonso and Flanagan’s (1999) criteria for evaluating the psychometric properties of early childhood diagnostic measures will provide criterion for evaluation of the adequacy of the data reviewed.

The DOCS is a norm-referenced instrument developed in an effort to establish a comprehensive screening instrument for young children who are at-risk. It purports to be an easy, clear, and quick method for screening the developmental status of children from birth to six-years-old. Parents or caregivers complete the questions on the DOCS, unless they are unable to read, in which case a professional should give assistance. The usefulness of the DOCS is in its ability to assess if children are developing typically, by making comparisons with other children of a similar age. There are three components to the DOCS. The first component is the Developmental Checklist (DC). The DC is a questionnaire that a parent of the child completes. It asks questions to establish the developmental level of children in four domains: cognition, language, social, and motor. The second component is the Adjustment Behavior Checklist (ABC). This section checks for any problems with children’s abilities to adjust and adapt to their environments. Part three is the Parental Stress and Support Checklist (PSSC). The PSSC
assesses the level of parental stress that is present in the family. It also examines how the 
stress and support of the family are influencing child development. Standard scores are 
available for the DC domains (Cognition, Motor, Social, and Language), the ABC, and 
the PSSC.

The publishers of the DOCS standardized the screener on 1,094 children from 30 
states. The demographic characteristics used for stratification include race, ethnicity, 
gender, urban/rural residence, and geographical region of the United States (e.g., 
Southeast, etc.). The stratification variables proportionately matched those found in the 
One exception would be a slight under representation in the “Other” ethnicity category 
(80% versus 89%). Variables reported but not controlled include parental age, gender, 
and education, and teacher age, gender, race, education, and experience.

Reiability. The publishers established reliability through a variety of methods. 
Coefficient alphas establish the internal consistency of items. The overall coefficient 
alphas for the DC domains were significant and strong for all of the age groups, with 
each one being either .98 or .99. The coefficient alphas for the ABC ranged from .81 to 
.90, but were considerably lower for the birth to 20 months age range, at .81. The upper 
age groups (over 20 months) were all stronger, with coefficient alphas between .88 and 
.90. Coefficient alphas for the PSSC, with all age ranges, were between .90 and .94. All 
of these reliabilities are adequate or good using Alfonso & Flanagan's (1999) criteria for 
evaluating preschool measures.

Test-retest reliability was established over 14-21 days to determine the stability of 
the test over time (n=98). The three groups used to establish test-retest reliability [one
group 2 to 3 years (n=35), and two groups 4 to 5 years) came from a restricted geographic region (Dallas and Sanger, Texas). The test-retest reliabilities for the DC overall domain were significant and yielded strong correlations between .91 and .96, for each of the age ranges. Test-retest reliability coefficients in this range are good, utilizing Alfonso and Flanagan’s criteria for evaluating adequacy of preschool measures. However, the DC Motor Domain correlation is slightly weaker for the early ages (2 to 3), with a significant test-retest reliability correlation coefficient of .85. Yet, this level is still good when compared to Alfonso and Flanagan’s (1999) recommendations. All of the age ranges for the DC Cognition Domain demonstrated weak correlations, between .85 and .88. The test-retest reliability correlations for the ABC were all significant and either .93 or .94. For the PSSC, the test-retest reliabilities were significant with correlations between .89 and .91. The DOCS exhibited adequate test-retest reliability when evaluated using Alfonso and Flanagan’s (1999) criteria. However, the publishers of the DOCS established the test-retest reliability on a relatively small and limited sample, making the results less certain.

The developers also reported interrater reliability. The overall interrater reliability correlation coefficient was .94. This interrater reliability came from a single study of parent and caregiver ratings of 30 children between the ages of 4 years, 0 months, to 5 years, 0 months. These children were from a restricted geographic region (Dallas, Texas).

Validity. The authors provided evidence for content validity through item development, item analysis, and an evaluation of item bias. The authors reviewed literature to establish item formats, resulting in two: parent/caregiver report and examiner
observation (when needed). Then, the publishers examined and analyzed other similar tests for young children, to evaluate whether the item content on the DOCS was representative of the developmental range. They conducted an item analysis to assess item difficulty, to ensure an appropriate floor and ceiling. The authors typically used items that fell between a difficulty level of .3 and .7, although there were some exceptions to this acknowledged. For the ABC and the PSSC, the publishers eliminated any items observed in over half of the parents and children, respectively, since that could not be an abnormal behavior. The DOCS authors also considered item discrimination when evaluating item analysis. They utilized the point-biserial correlation technique to determine item discrimination. Generally, the publishers accepted items if the item discrimination coefficients were significant and the magnitudes were between .30 and .80. They also examined item bias, eliminating any items that had difficulty differences due to race and gender that were greater than .10. The publishers analyzed the items multiple times. They purport that the final analysis supported the appropriateness of the remaining items. The process and criteria used for determining item difficulty are customary for test development. These results are acceptable when using Alfonso and Flanagan’s (1999) criteria.

The DOCS publishers established criterion-related validity with many different measures, using them as criteria for comparisons. For example, they used the Bayley Scales of Infant Development (BSID; Bayley, 1969) for comparison with the DC domains. The Motor Domain of the DC when compared to the BSID revealed a correlation of .45. The correlation between the Mental component of the BSID and the overall DC component of the DOCS was also .45. In fact, all of the reported correlations
were significant, beyond the .01 level. One would expect these correlations to be low to moderate, since a screening instrument, by its very nature, should not yield the same results as a diagnostic assessment tool.

Construct validity was the last evidence of validity provided by the authors. Since the DOCS is a test that yields a developmental level, the authors pointed out that DC performance should be higher for children that are older and have more experience. The overall age correlation was significant and strong ($r = .83$). The authors also stated that because there was measurement overlap across the DC domains, significant and positive correlations should emerge. The DC domains all had significant and strong correlations with each other ranging from .97 to .99. The authors also predicted a relationship between the ABC and PSSC scores, since children’s overall functioning affects the PSSC. The ABC and the PSSC had a significant correlation of .44, demonstrating the predicted connectedness. The authors also purported that the DC component should correlate with intelligence. The Verbal Reasoning section of the Stanford Binet, 4th Edition (SB-4; Thorndike, Hagen, & Sattler, 1986) and the Language domain of the DC component of the DOCS had a correlation of .46. When the authors compared the same section of the DOCS to the Quantitative Reasoning section of the SB-4, the correlation was .45. All comparisons were statistically significant indicating a significant relationship between the DC component of the DOCS and the SB-4. Finally, in assessing construct validity, the authors suggested that, since the DC measures natural and acquired abilities, there should be a distinction among the DC subtests between children who are developing typically and atypically (e.g., children with deficits in cognition, language, social skills, etc.). Some of the correlations between the diagnoses (language impaired,
etc.) and the DC overall scores were strong, but others were weak. The overall score correlation for the early childhood intervention group was a modest .78 (domains ranged from .75 to .79). The correlation between the substance exposed group and the overall DC score was only a .82 (domains ranged from .74 to .83). The correlation for the overall DC score with two groups of at-risk students (identified due to placement in preschool intervention or a private school for children exhibiting early learning problems) was .87 and .95. The domain intracorrelations ranged from .82 to .90. The authors used these numbers to display the DOCS capacity to provide a measure that differentiates population groups. The reported psychometric properties for construct validity are considered acceptable to good using Alfonso and Flanagan’s (1999) criteria for preschool measurement instruments. It is important to note that the DOCS publishers review the entire DOCS, while this study focuses solely on the DC, since the DC is the common choice in practice.

Overall, Hresko et al. (1994) provide a wide range of evidence for reliability and validity. However, the evidence is not conclusive. For example, there were not many diagnostic tests utilized for comparative purposes to ascertain construct validity. Sample sizes and age ranges were limited in the test-retest and interrater reliability studies. In addition, the authors provided limited evidence for the birth to three-age range, yet this age group must receive the DC in Kentucky. In addition, all of these studies come from the authors of the DOCS without independent verification. Although the instrument was published in 1994, there has been no further research published about the psychometric properties of the DOCS. A search of the Ebscohost computer database, on February 17, 2006, resulted in five hits. The key words/phrases used were “Developmental
Observation Checklist System,” and “DOCS.” Five studies utilized the DOCS to describe population samples and did not investigate psychometric properties. The use of the DOCS in such studies provides some support for the DOCS as a reputable measurement tool and indicates a respect for the scores it provides. However, they do not provide data to support the psychometric adequacy of the measure. Even more important are data to support its use as a screening instrument including such properties as specificity, sensitivity, and classification consistency, as noted by Glascoe (1991). Glascoe’s review of issues in developmental screening recommends the investigation of sensitivity, specificity and classification consistency with screening measures. Further, he advocates viewing sensitivity as good if it is above 80%, specificity at 90%, and classification consistency of at least 80%.

**Summary and Purpose**

This investigation focuses on the validity of the Developmental Checklist (DC) of the DOCS as a screening tool for identification of children in need of early intervention. Experts in the field and the research reviewed indicate that early intervention is important because it can produce many short-term and long-term benefits for children, families, and society. Such benefits include academic or cognitive gains, improved high school graduation rates, greater success in the labor market, behavioral benefits, decreases in future crime rates, diminished amounts of hospitalizations and injuries, and overall economic gains for society as a whole. Since early intervention can have such a positive impact on society, it is important to identify children in need of services in a timely manner. The main purpose of early intervention is to provide interventions for children with special needs in order to help them acquire the skills they need to be as independent
as possible in as many environments as possible (Chen, 1999). Thus, it is critical to evaluate the tools used in the referral process for early intervention programs.

In addition, both federal legislation and Kentucky Revised Statutes (KRS, Kentucky Revised Statutes, 2006) provide guidelines for services for infants and toddlers with disabilities and developmental delays. Statute 200.650 states that there is an urgent need for early intervention to enhance the development of infants and toddlers in an effort to increase their future independence and decrease future delay. Statute 200.650 also states that proper early identification is important because it will reduce future family, school, and social services costs, by avoiding such results as special education and institutionalization. In addition, Kentucky statute 200.668 instructs the Cabinet for Health and Family Services to implement child find systems in an effort to assist (and hence place) all eligible children. Further, Kentucky legislation mandated the use of the DOCS in the assessment process to determine services for KEIS – First Steps (Kentucky Cabinet for Health and Family Services, 2006a). It is imperative to evaluate the tools used to determine if they are adequate. Appropriate early interventions that will maximize the future successes of young children can only be possible if the tools used for early identification are valid and accurate.

The DOCS is the required screening tool in Kentucky’s early intervention system. However, there is little independent verification of the author’s claims of adequate or good psychometric properties and no evidence of specificity, sensitivity, and classification consistency. Additionally, the population data are from the normative samples (a population of individuals with no identified special needs). In contrast, professionals utilize the instrument for a referred population, comprised of children with
questionable developmental status. It is also important to note that many professionals administer the DC of the DOCS via an interview, which is inconsistent with the standardization of the DOCS (in which parents filled out the protocols independently).

The present investigation will explore the usefulness of the Developmental Checklist (DC) as a screening instrument. It is important to understand the psychometric properties of the DOCS and its relationship with one of the most commonly used diagnostic measures, the Battelle Developmental Inventory, 2nd Edition (BDI-2). The BDI-2 is the criterion measure in this study as it is a frequently used diagnostic measure employed in the Primary Level Evaluation. The Developmental Checklist (DC) component of the DOCS has four domains, which roughly correspond with four of the five BDI-2 domains (cognition: cognitive, language: communication, social: personal-social, and motor: motor). In order to ascertain if the DC is a useful instrument for screening purposes, it is necessary to determine if a significant relationship exists and the nature of the relationship between the DC and the BDI-2. If there is a significant relationship between the DOCS and the diagnostic measure, then a positive correlation that is moderate to strong should emerge. This should be evident for the total score and the paired domains.

A second purpose for the current study is to examine classification consistency between the DOCS DC and a diagnostic measure, the BDI-2. Sensitivity refers to how accurately the DOCS DC identifies children as needing further evaluation, when a criterion measure (BDI-2) indicates a need for further evaluation (true positives). Specificity refers to how accurately the DOCS DC identifies children as not being in need of further evaluation, as revealed by the BDI-2 (true negatives). Ideally, there will be a
significantly high rate of true positives and true negatives (high classification consistency). However, disagreements between the tests will emerge. When this occurs, it is preferable that the DOCS DC be too sensitive (identifying children as at-risk when they actually are not) than not sensitive enough (claiming children are not at-risk, when they actually are). It is far less of a problem to test too much, than to leave testing undone, with children unidentified and not served. The present study will examine how consistently the DOCS DC classifies children as at-risk or not at-risk when compared to the BDI-2. Classification consistency or consistency between the two measurements will yield information about the usefulness of the DOCS DC as a screener. Specific hypotheses for this investigation will be as follows:

Hypothesis One. (a) The domains of the DC and the BDI-2 will evidence significant and high correlations ($r \geq .60$) between like domains and total scores. Similar domain comparisons are (DC/BDI-2): Language/Communication, Social/Personal-Social, Social/Adaptive, Motor/Motor, and Cognition/Cognitive. Total score comparisons are (DC/BDI-2): Language/Total Score, Social/Total Score, Motor/Total Score, and Cognition/Total Score.

(b) The dissimilar domains of the DC and the BDI-2 will evidence low to moderate correlations ($r = .40$ to $.59$). Dissimilar domain comparisons are (DC/BDI-2): Language/Personal-Social, Adaptive, Motor, and Cognitive; Social/Communication, Motor, and Cognitive; Motor/Communication, Personal-Social, Adaptive, and Cognitive; Cognition/Communication, Personal-Social, Adaptive, and Motor.
Hypothesis Two. High classification consistency (≥ 80%) will be evidenced between the domains of the DC and the BDI-2 (using the Kentucky classification cutoff of 2 SD on one domain or 1½ SD on two domains).
Method

Subjects

The present research utilized archival study records from one First Steps Point of Entry (POE) program in Kentucky to obtain the sample. A Point of Entry staff member identified 51 records of young children evaluated between May 2005 and May 2007 with a Primary Level Evaluation (PLE) report on file containing the DC and the BDI-2. The individuals within this sample were between the ages of 2 and 35 months at the time of the PLE report with a mean age of 19 months. The sample consisted of 43 Caucasians (84%), 3 African Americans (6%), 1 Asian (2%), 3 Hispanics (6%), and 1 unknown (2%, the information was not available). Thirty-six males (71%) and 15 females (29%) comprised the sample. POE staff recorded financial income in two different ways. A large portion of the sample came from families whose income was below 200% of poverty level, which made them eligible for a Kentucky subsidized medical card. Families can also be eligible for a medical card if a child suffers from a permanent disability or an extreme medical situation, which is above that family’s financial means, according to KEIS Evaluation Specialist N. Slaughter (personal communication, June 8, 2007). If a family was eligible for a medical card, the POE staff did not include family income, other than to indicate the presence of a medical card. If a family was not eligible for the medical card, then the POE staff reported their actual family income. Thirty-five of the individuals within the sample (69%) possessed a Kentucky subsidized medical card. Sixteen (31%) had reported salaries ranging from $27,174 to $127,000, with a mean salary of $53,098.67.
**Instruments**

*Developmental Observation Checklist System (DOCS; Hresko et al., 1994)*

*Developmental Checklist (DC)*. The DOCS is a screening instrument used to identify children from birth to six years who are at-risk. Practitioners compare referred children’s scores with other children’s scores of similar age from the normative sample. The publishers of the DOCS established reliability through a variety of methods. They utilized coefficient alphas to establish internal consistency, with the overall coefficient alphas for the DC domains being either .98 or .99. The test-retest reliabilities reported were all adequate, each being at least .85. The publishers also reported interrater reliability, with the overall interrater reliability correlation coefficient being .94. These reliabilities are adequate to good using Alfonso and Flanagan’s (1999) criteria for evaluating preschool measures.

The publishers also provided validity studies. They analyzed item difficulties and typically used items that had a difficulty level between .3 and .7 (items not too easy or too hard, so they would discriminate). Criterion-related validity was also determined through concurrent procedures. For example, the motor domains of the DOCS and the Bayley Scales of Infant Development (BSID) correlated at .45. The publishers also provided construct validity. The overall age correlation was significant and strong ($r = .83$) although some weaknesses do exist (e.g., low numbers in certain age groups). Overall, the reliabilities and validities provided by the DOCS publishers are good using Alfonso and Flanagan’s (1999) criteria. However, the authors did not provide evidence for validity with a referred population, which is the intended use. In addition, there is no information regarding sensitivity, specificity, and classification consistency. It is
important to note that the DOCS publishers review the entire DOCS, while this study
focuses solely on the DC, since the DC is what is commonly used in practice.

_Battelle Developmental Inventory, 2nd Edition (BDI-2; Newborg, 2005)._ This
instrument will be used as the criterion measure against which the DC results are
compared. The BDI-2 is a comprehensive assessment tool utilized to measure the
development of children from birth to eight years. The publishers of the BDI-2 created
the second edition to update the materials and norms and to make it more user friendly.
As reported previously, it is a commonly used measure in Kentucky because it measures
each of the five federally mandated developmental domains required for determination of
eligibility for early intervention.

The publishers of the BDI-2 established reliability in several different ways. The
internal consistency was shown to be high ($r = .99$). Test-retest reliability was $.93$ and
$.94$ for the two-year-old group and four-year-old group, respectively. The inter-scorer
agreement on items was between $.94$ and $.99$. These results are indicative of good
reliability according to Alfonso and Flanagan’s (1999) criteria.

The publishers of the BDI-2 also established validity in a variety of ways. The
BDI-2 displayed lower divergent validities with the BDI than convergent validities (.32
to .72 versus .64 to .78). BDI-2 comparisons with other assessment tools, like the Bayley
Scales of Infant Development, Second Edition (BSID-2), rendered similar results
(Bayley, 1993). Construct validity was also provided, with $90\%$ of the correlations
between domain scores across age ranges being significant (between $.3$ and $.7$).
Empirical item analysis also evidenced content validity. The publishers typically
eliminated items with poor Rasch-fit statistics from the BDI-2. Overall, the psychometric
properties of the BDI-2 were good according to the criteria for preschool measures set by Alfonso and Flanagan (1999).

Procedure

To comply with the Health Insurance Portability and Accountability Act (HIPAA) privacy of records regulations, a Point of Entry clerical staff member received a stipend to retrieve data from the POE files. The staff member transferred data onto a demographic/test score form used to record DC and BDI-2 scores (see Appendix A). The POE staff member assigned a code to each form and created a master list of all of the codes along with an ID sheet. The POE staff kept the code sheet to verify scores in case of transcription errors. The WKU Institutional Review Board for Human Subjects and the Kentucky Cabinet for Health and Family Services Institutional Review Board approved all of the procedures (see Appendix B).
Results

The analysis of the data was descriptive in nature. Table 1 provides the descriptive test data from the DC of the DOCS and the criterion measure (BDI-2; means, standard deviation, range). The mean standard score for the sample for each of the domains on each measure were all low when compared to a normative mean of 100. The standard score means for the BDI-2 ranged from 72.75 to 87.31, whereas the DC means ranged from 87.04 to 88.80. All BDI-2 domain means were at least one SD below the mean except for the Motor Domain. Overall, the DC domain mean standard scores were within the Below Average classification range (almost Average) while the BDI-2 mean domain standard scores were within the Low Average to Borderline classification range.

To address the first hypothesis (significant and high correlation between like domains and total scores on the DC and the BDI-2 will be found and a low to moderate correlation on dissimilar domains) a correlation matrix was calculated using the Pearson r statistic between domains on the DC and the BDI-2 and the BDI-2 Total Score (Table 2). The Bonferroni approach to control for Type 1 errors across the 20 correlations established a p value of less than .003 (.05/20 = .0035) for significance. The present research did not find support for the first part of Hypothesis One. No significant, high (above r = .60) correlations were found between any of the like domains (Language, Motor, Cognition and Social) for the DC with the BDI-2. Comparisons could not be made with the DC Total Score, because that information was not available.

The present research did not find support for the second part of Hypothesis One. This study found no significant correlations between domains across measures. However, the present research did find significant intracorrelations (correlations between the domains of the same measure). Each of the domains of the DC were significantly
Table 1

*Descriptive Statistics for the Developmental Checklist (DC) of the Developmental Observation Checklist System (DOCS) and the Battelle Developmental Inventory, 2nd Edition (BDI-2)*

<table>
<thead>
<tr>
<th>Measure and Domain</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>DC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Language</td>
<td>87.04</td>
<td>10.61</td>
<td>28 – 107</td>
</tr>
<tr>
<td>Motor</td>
<td>87.86</td>
<td>8.24</td>
<td>64 – 108</td>
</tr>
<tr>
<td>Social</td>
<td>88.80</td>
<td>7.73</td>
<td>64 – 108</td>
</tr>
<tr>
<td>Cognition</td>
<td>87.69</td>
<td>7.17</td>
<td>64 – 107</td>
</tr>
<tr>
<td>BDI-2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adaptive</td>
<td>82.14</td>
<td>14.75</td>
<td>55 – 104</td>
</tr>
<tr>
<td>Communication</td>
<td>72.75</td>
<td>13.46</td>
<td>55 – 106</td>
</tr>
<tr>
<td>Motor</td>
<td>87.31</td>
<td>17.80</td>
<td>55 – 129</td>
</tr>
<tr>
<td>Cognitive</td>
<td>77.88</td>
<td>13.69</td>
<td>57 – 110</td>
</tr>
<tr>
<td>Personal-Social</td>
<td>80.47</td>
<td>10.92</td>
<td>59 – 103</td>
</tr>
<tr>
<td>Total</td>
<td>75.57</td>
<td>13.37</td>
<td>51 – 103</td>
</tr>
</tbody>
</table>

*Note.* N=51.
Table 1

Correlations of Developmental Checklist (DC) of the Developmental Observation Checklist System (DOCS) Domains with Battelle Developmental Inventory, 2nd Edition (BDI-2) Domains and Total Score

<table>
<thead>
<tr>
<th>Domain</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>DL</td>
<td></td>
<td>.817*</td>
<td>.853*</td>
<td>.864*</td>
<td>.134</td>
<td>.289</td>
<td>.067</td>
<td>.128</td>
<td>.189</td>
<td>.206</td>
</tr>
<tr>
<td>DM</td>
<td></td>
<td></td>
<td>.932*</td>
<td>.874*</td>
<td>.320</td>
<td>.371</td>
<td>.303</td>
<td>.247</td>
<td>.289</td>
<td>.395</td>
</tr>
<tr>
<td>DS</td>
<td></td>
<td></td>
<td></td>
<td>.909*</td>
<td>.291</td>
<td>.379</td>
<td>.215</td>
<td>.291</td>
<td>.258</td>
<td>.366</td>
</tr>
<tr>
<td>DC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.186</td>
<td>.316</td>
<td>.104</td>
<td>.193</td>
<td>.225</td>
<td>.260</td>
</tr>
<tr>
<td>BDI-A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.268</td>
<td>.538*</td>
<td>.479*</td>
<td>.586*</td>
<td>.704*</td>
</tr>
<tr>
<td>BDI-CM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.372</td>
<td>.435*</td>
<td>.552*</td>
<td>.687*</td>
</tr>
<tr>
<td>BDI-M</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.543*</td>
<td>.575*</td>
<td>.824*</td>
</tr>
<tr>
<td>BDI-CG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.564*</td>
<td>.780*</td>
</tr>
<tr>
<td>BDI-PS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.835*</td>
</tr>
<tr>
<td>BDI-T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


* p < .003
correlated at a strong level ranging from .817 to .932. All but two of the BDI-2 domains were significantly correlated with one another at a moderate to strong level (.435 to .835). The nonsignificant BDI-2 correlations were for the Communication domain when correlated with the Adaptive and Motor domains.

The second hypothesis predicted high classification consistency between the DC and the BDI-2. Tables 3 and 4 contain the results for the contingency tables for each of the like domains across measures using the BDI-2 domain score as the criterion. Table 3 contains the contingency tables using the 2 SD criterion on the BDI-2 domains (as the placement criteria). Table 4 contains the contingency table using the 1½ SD on two domains of the BDI-2 (as the placement criteria). Tables 3 and 4 also reflect various "cutoff" levels on the DC. The state recommends a standard score of 94 and below for referral for further evaluation (Kentucky Cabinet for Health and Family Services, 2006a). The cutoff scores of 85 (1 SD below the mean) and 77 (1½ SD below the mean) are for comparison purposes. A standard score of 70 (2 SD below the mean) was deemed unnecessary, as the pattern of information became redundant (see Discussion). The DC contains four domains, each one roughly corresponding with one of the five domains from the BDI-2. This research paired the fifth domain from the BDI-2 (Adaptive) with the DC Social, once again, since it is the closest match and contains the most skill overlap. Sensitivity (true positives/correct identifications) and specificity (true negatives/correct dismissals) were added to obtain the classification consistency. Classification consistency of over .80 was the criterion set for good agreement.

Hypothesis Two predicted high classification consistency between the domains of the DC and the BDI-2. Three comparisons of domain classifications between the two measures supported this hypothesis. Utilizing the 2 SD criterion on the BDI-2 (Table 3),
Table 3

*Classification Agreement for the Developmental Checklist (DC) Domains with Battelle Developmental Inventory, 2nd Edition (BDI-2) Domains at 2 Standard Deviation (SD) Criterion*

<table>
<thead>
<tr>
<th>Domain</th>
<th>Cutoff</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Consistency</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Communication/Language (BDI-2/DC)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>94</td>
<td>91%</td>
<td>14%</td>
<td>49%</td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>61%</td>
<td>79%</td>
<td>71%</td>
<td></td>
</tr>
<tr>
<td>77</td>
<td>9%</td>
<td>100%</td>
<td>59%</td>
<td></td>
</tr>
<tr>
<td><strong>Motor (BDI2/DC)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>94</td>
<td>80%</td>
<td>15%</td>
<td>27%</td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>70%</td>
<td>76%</td>
<td>75%</td>
<td></td>
</tr>
<tr>
<td>77</td>
<td>10%</td>
<td>95%</td>
<td>78%</td>
<td></td>
</tr>
<tr>
<td><strong>Personal-Social/Social (BDI-2/DC)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>94</td>
<td>83%</td>
<td>18%</td>
<td>25%</td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>50%</td>
<td>73%</td>
<td>71%</td>
<td></td>
</tr>
<tr>
<td>77</td>
<td>17%</td>
<td>96%</td>
<td>86%</td>
<td></td>
</tr>
<tr>
<td>77</td>
<td>18%</td>
<td>98%</td>
<td>80%</td>
<td></td>
</tr>
<tr>
<td>94</td>
<td>91%</td>
<td>10%</td>
<td>27%</td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>64%</td>
<td>78%</td>
<td>75%</td>
<td></td>
</tr>
</tbody>
</table>
Table 3 (continued).

<table>
<thead>
<tr>
<th>Domain</th>
<th>Cutoff</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Consistency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adaptive/Social (BDI-2/DC)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>94</td>
<td>90%</td>
<td>20%</td>
<td>33%</td>
</tr>
<tr>
<td></td>
<td>85</td>
<td>80%</td>
<td>83%</td>
<td>82%</td>
</tr>
<tr>
<td></td>
<td>77</td>
<td>10%</td>
<td>95%</td>
<td>78%</td>
</tr>
</tbody>
</table>

*Note.* Cutoff scores are standard scores. 2 SD = standard score \( \leq 70 \).
Table 4

Classification Agreement for the Developmental Checklist (DC) Domains with Battelle Developmental Inventory, 2nd Edition (BDI-2) Domains at 1½ Standard Deviation (SD) Criterion

<table>
<thead>
<tr>
<th>Domain</th>
<th>Cutoff</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Consistency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communication/Language (BDI-2/DC)</td>
<td>94</td>
<td>88%</td>
<td>12%</td>
<td>63%</td>
</tr>
<tr>
<td></td>
<td>85</td>
<td>53%</td>
<td>88%</td>
<td>65%</td>
</tr>
<tr>
<td></td>
<td>77</td>
<td>6%</td>
<td>100%</td>
<td>37%</td>
</tr>
<tr>
<td>Motor (BDI-2/DC)</td>
<td>94</td>
<td>88%</td>
<td>18%</td>
<td>41%</td>
</tr>
<tr>
<td></td>
<td>85</td>
<td>59%</td>
<td>79%</td>
<td>73%</td>
</tr>
<tr>
<td></td>
<td>77</td>
<td>18%</td>
<td>100%</td>
<td>73%</td>
</tr>
<tr>
<td>Personal-Social/Social (BDI-2/DC)</td>
<td>94</td>
<td>83%</td>
<td>18%</td>
<td>47%</td>
</tr>
<tr>
<td></td>
<td>85</td>
<td>39%</td>
<td>79%</td>
<td>61%</td>
</tr>
<tr>
<td></td>
<td>77</td>
<td>9%</td>
<td>96%</td>
<td>57%</td>
</tr>
<tr>
<td>Domain Cutoff</td>
<td>Sensitivity</td>
<td>Specificity</td>
<td>Consistency</td>
<td></td>
</tr>
<tr>
<td>---------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>Cognitive/Cognition (BDI-2/DC)</td>
<td>94</td>
<td>84%</td>
<td>0%</td>
<td>51%</td>
</tr>
<tr>
<td></td>
<td>85</td>
<td>39%</td>
<td>80%</td>
<td>55%</td>
</tr>
<tr>
<td></td>
<td>77</td>
<td>10%</td>
<td>100%</td>
<td>45%</td>
</tr>
<tr>
<td>Table 4 (continued).</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Domain</td>
<td>Cutoff</td>
<td>Sensitivity</td>
<td>Specificity</td>
<td>Consistency</td>
</tr>
<tr>
<td>Adaptive/Social (BDI-2/DC)</td>
<td>94</td>
<td>89%</td>
<td>21%</td>
<td>45%</td>
</tr>
<tr>
<td></td>
<td>85</td>
<td>50%</td>
<td>82%</td>
<td>71%</td>
</tr>
<tr>
<td></td>
<td>77</td>
<td>6%</td>
<td>94%</td>
<td>63%</td>
</tr>
</tbody>
</table>

*Note.* Cutoff scores are standard scores. $1\frac{1}{2}$ SD = standard score $\leq 77$. 
only three comparisons reached the 80% level of classification consistency. These were the Personal-Social and Social comparison at the DC cutoff of 77, Cognitive/Cognition comparison at the DC cutoff of 77, and the Adaptive/Social comparison at the DC cutoff of 85. Overall, the present research found limited evidence for good classification consistency for comparisons made with the BDI-2 at two standard deviations below the mean. Only three out of 15 comparisons evidenced good consistency (>80%). At the 1½ standard deviation criterion on the BDI-2 (Table 4), the present research did not find good consistency at any of the DC levels for any of the domain comparisons. The 1½ SD criterion provides no support for classification consistency between the two measures.

Sensitivity (true positives/correct identifications) and specificity (true negatives/correct dismissals) were computed to further understand the results. Glascoe (1991) recommends a level of 80% for sensitivity and 90% for specificity for screening measures. Using both the 1½ and the 2 SD criterions for placement on the BDI-2 (Tables 3 and 4), sensitivity was at or above 80% for each of the domain comparisons with the cutoff standard score of 94 and below on the DC. In addition, the Adaptive/Social comparison for the 2 SD criterion for placement on the BDI-2 (Table 3) with a DC cutoff score of 85 also evidenced good consistency. Specificity of 90%, the level recommended by Glascoe as desirable for screening measures, was reached for all domain pairs only at the most stringent DC cutoff (standard score of 77) at both the 1½ and 2 SD placement criterion levels on the BDI-2 (Tables 3 and 4).
This researcher conducted actual placement eligibility status for the sample by examining the BDI-2 and DC scores across domains. The placement status was determined by using the 1½ SD’s on two domains or 2 SD’s on one domain for the BDI-2, as dictated by Kentucky standards for placement, against scores of ≤ 94 on the DC, as dictated by the Kentucky Administration Regulations. Overall, 48 of the 51 cases were established as being at-risk by the DC (cases with scores ≤ 94), while the BDI-2 established 36 of these DC cases as being truly eligible for services (75%). In addition, of the remaining three cases from the DC that were not identified as being at-risk (cases without any scores ≤ 94) the BDI-2 identified each one of these cases as being eligible for services. The present researcher obtained this information by checking the individual cases with their ID numbers.
Discussion

The purpose of the present investigation was to evaluate the DC component of the DOCS to provide evidence for the use of the DC as an adequate and appropriate screening instrument for use with referred children. The present investigation hypothesized that there would be significant and high correlations for the like domains and total scores on the DC and the BDI-2. Although correlations would not be expected to approach 1.0 (the DOCS is a screener and not a comprehensive assessment tool), a strong correlation of greater than .60 would be expected, which would support the use of the DC as a screening tool. Domains that are not strongly correlated (less than .60) would call into question whether the two measures are measuring the same constructs or if one measure is doing so to a higher degree. It also questions whether practitioners should use the measures together. The second part of this hypothesis predicted that there would be a low to moderate correlation (.40 - .59) between dissimilar domains. For example, one would expect the Language domain of the DC to have a low to moderate size correlation to the Motor domain of the BDI-2, not a strong correlation ($r > .60$), since the Language and Motor domains should not be measuring the same constructs. Consequently, if a strong correlation emerged, then the underlying measured constructs would become suspicious, making one question if the underlying constructs assessed by each instrument were in fact different. If there were no significant relationship, one would again question the legitimacy of the DC and its use with the BDI-2.

To address the first purpose, of determining a relationship between the DC and the BDI-2, this researcher calculated a correlation matrix using the Pearson $r$ statistic between domains on the DC and the BDI-2 and the BDI-2 Total Score. The correlations
between like domains for the two measures were not significant. These findings did not support the first hypothesis (significant and strong correlations between like domains). The BDI-2 is a new measure; therefore, other correlation studies were not available between the DC and the BDI-2. However, the publishers of the DOCS reported correlations between the components of the DC with other well-known norm-referenced diagnostic assessment tools. For example, the DC publishers report a significant correlation between the DC Language Domain and the McCarthy Scales of Children’s Abilities Verbal Domain of .67. Additionally, the publishers of the DC also report a correlation of .57 between the McCarthy Motor Domain and the DC Motor Domain. Overall, the DOCS publishers report significant criterion-related validity correlations at the moderate to high level, while the present research found none. Also noteworthy, are the small sample sizes of the reported DOCS correlation studies (20 to 35). If significant correlations are not present between the DC and BDI-2, questions emerge regarding whether the use of these measures together is advantageous.

From the present study, the strongest correlations (correlations ≥ .60) were between domains from the same assessment tool (e.g., DC Language and DC Motor, BDI-2 Language and BDI-2 Cognition, etc.) as well as comparisons between BDI-2 domains and the BDI-2 Total Score (BDI-2 Motor and BDI-2 Total Score, etc.). The intracorrelations reported in the DOCS manual range from .97 to .99 and are stronger than the intracorrelations found in the present study (.82 to .93).

The second part of the first hypothesis predicted low to moderate correlations (.40 to .59) between dissimilar domains across measures. The publishers of the DOCS report correlations between the DC and other dissimilar domains from other measures, while
this study did not find any significant correlations between dissimilar domains. The correlations between dissimilar domains reported by the DC publishers (with McCarthy, SB-4, etc.) were .35 to .71. Overall, the DOCS publishers reported several criterion-related validity correlations that were at the low to moderate level for dissimilar domains, while the present research did not find any. The present correlations are also low in comparison to the atypical group (groups of at-risk children) intracorrelations reported in the DC manual (.82 to .95) as well as the early childhood intervention group intracorrelations (.75 to .79). Although these correlations are more modest than the other DOCS reported correlations, they are stronger than the intracorrelations from the present study (comprised of a referred population).

The second purpose of the study was to determine whether there is classification consistency between the DC of the DOCS and the BDI-2 (using Kentucky classifications). It was the expectation of this researcher that there would be high agreement between the two measures (.80 or higher). Assuming there was high classification consistency, this would support the use of the DC at KEIS – First Steps programs. If the agreement between the two measures was not high and the data demonstrated that the DC was too sensitive (identifying too many children as at-risk when they truly are not) then the DC would still be an adequate measure to use, albeit an expensive choice. However, if the DC was shown to have poor sensitivity rates (missing a lot of children who need further testing), then major questions would emerge with regard to whether or not it should be used as the exclusive screening instrument for the Kentucky First Steps programs.
Of the 30 contingency tables calculated, the present research only found classification consistency in three cases. As would be expected, sensitivity improved within each domain as the DC “at-risk” criteria became less stringent (standard scores of 77, 85, and 94). However, this was not true of the classification consistency comparisons. Moreover, although sensitivity was highest with a standard score of 94 on the DC, specificity was typically the weakest. Additionally, over-identification became a more frequent problem. Thus, classification consistency may have become poorer, due to an increased level of false identifications (recall, the DC identified 48 children as “at-risk” at a standard score of 94 on the DC, while the BDI-2 only identified 36 of those children as eligible for services).

Limitations

The present research provides important information; however, there are several limitations to this study. For example, the sample size for this study is relatively small (n = 51). An additional threat to external validity is that the POE staff collected the data from a restricted geographical region. The sample was from one KEIS-First Steps POE program in Kentucky. In addition, because the sample is a referred population, with a very low ratio of typically developing children, specificity is hard to accurately determine. The proportion of minority participation of 16% is somewhat higher than the state ratio (9%), yet slightly lower when compared to the national minority population of 23% (U.S. Census Bureau, 2005). Another potential threat is from the testing method utilized. A parent interview violates the standardization of the DOCS that prescribes for the parents independently complete the form. Also, the state mandates that the DOCS be given (Kentucky Cabinet for Health and Family Services, 2006a), while in practice, it is
typically just the DC. Internal threats to the validity of the study may also be present. Maturation between testing sessions (from the screening to the comprehensive testing) may have occurred, possibly affecting the test results.

**Implications**

**Practical implications.** The findings from this study have important implications for practitioners in Kentucky’s First Steps Early Intervention program. Due to the critical nature of early intervention, it is vital that the screening processes be psychometrically sound in order to identify the need for intervention. The present study provides relevant information regarding the properties of the DC and the use of the DC with the BDI-2. This study did not find support for the use of the DC with the BDI-2. While limitations exist, the data set is quite reasonable for a referred sample. In addition, the study data comes from First Steps program records, instead of contrived or manufactured data. Additionally, this study is important because it provides statistical data that are independent from the publishers of the DC.

When legislative rationale is adhered, and scores at or below 94 regarded as at-risk, then the DC becomes far more sensitive to those children with potential disabilities. However, a tendency to over-refer often occurs, leading to wasted time and money. Yet, the problem of under-referral is not minimized or eliminated. Children are still missed who would actually qualify if given a comprehensive evaluation (recall the three children from the present study who were not at-risk, according to the DC, yet qualified on the BDI-2). Moreover, the whole purpose of a screener is to determine who needs further testing, or in other words, who is at-risk, and separate those children from those not at-risk, who do not need further evaluation (Nuttall, et al., 1999). If this is not the result,
then money, time, and effort are wasted. In the present research, the fact that there are no significant correlations elicits doubt regarding the legitimacy and predictive value of the DC as a screener when combined with the BDI-2.

Recommendations for future research. The current study investigated the predictive ability of the DC as a screener for early intervention programs. Future research may want to further examine the DC, along with the rest of the DOCS, to provide information about the predictive ability of the DOCS, as whole, as a screening instrument for identifying disabilities and/or developmental delays. Other professionals should conduct further research across larger geographic areas to enhance the ability of one to generalize the results. This could also lend itself to a larger sample. Additionally, other professionals should conduct studies that analyze the use of the DC with other norm-referenced diagnostic tools. Studies should also take place to tease out whether the lack of correlation between DC and the BDI-2 is attributable to the restricted sampling or other method anomalies. Finally, when the DOCS is utilized, practitioners should utilize the screener as mandated by the KRS, with a comprehensive follow-up assessment when any DOCS scores are 94 or below.
References


Appendix A

Demographic/Test Score Form
Demographic/Test Score Form

**From Intake 1:**

Gender: M____ F_____ Age at testing: ____________ months

Race: Caucasian _____ African American _____ Asian _____ Hispanic _____

**From First Steps Financial Information Form**

Most recent gross family income ______________________________________

Family’s income is below 200% of poverty level: Yes _____ No _____

**From test protocols**

<table>
<thead>
<tr>
<th>DOCS Domain</th>
<th>Component Quotient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Developmental Language Quotient (DLQ)</td>
<td></td>
</tr>
<tr>
<td>Developmental Motor Quotient (DMQ)</td>
<td></td>
</tr>
<tr>
<td>Developmental Social Quotient (DSQ)</td>
<td></td>
</tr>
<tr>
<td>Developmental Cognition Quotient (DCQ)</td>
<td></td>
</tr>
<tr>
<td><strong>Overall Developmental Quotient (ODQ)</strong></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BDI-2 Domain</th>
<th>Developmental Quotient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adaptive</td>
<td></td>
</tr>
<tr>
<td>Personal-Social</td>
<td></td>
</tr>
<tr>
<td>Communication</td>
<td></td>
</tr>
<tr>
<td>Motor</td>
<td></td>
</tr>
<tr>
<td>Cognitive</td>
<td></td>
</tr>
<tr>
<td><strong>BDI-2 Total</strong></td>
<td></td>
</tr>
</tbody>
</table>
Appendix B

Human Subjects Review Board Approval
Dear Monte:

Your revision to your research project, "Validity of the Developmental Checklist of the Developmental Observation Checklist System," was reviewed by the HSRB and it has been determined that risks to subjects are: (1) minimized and reasonable; and that (2) research procedures are consistent with a sound research design and do not expose the subjects to unnecessary risk. Reviewers determined that: (1) benefits to subjects are considered along with the importance of the topic and that outcomes are reasonable; (2) selection of subjects is equitable; and (3) the purposes of the research and the research setting is amenable to subjects' welfare and producing desired outcomes; that indications of coercion or prejudice are absent, and that participation is clearly voluntary.

1. In addition, the IRB found that you need to orient participants as follows: (1) signed informed consent is not required as data is being collected by a secondary data source; (2) Provision is made for collecting, using and storing data in a manner that protects the safety and privacy of the subjects and the confidentiality of the data. (3) Appropriate safeguards are included to protect the rights and welfare of the subjects.

This project is therefore approved at the Exempt Review Level until August 15, 2007 contingent upon IRB approval from the Cabinet for Health and Family Services.

2. Please note that the institution is not responsible for any actions regarding this protocol before approval. If you expand the project at a later date to use other instruments please re-apply. Copies of your request for human subjects review, your application, and this approval, are maintained in the Office of Sponsored Programs at the above address. Please report any changes to this approved protocol to this office.

Sincerely,

Sean Rubino, M.P.A.
Compliance Manager
Office of Sponsored Programs
May 7, 2007

Monte Gannon
1347 Bowie Trail, Apt A
Owensboro, Kentucky 42303

Re: CHFS IRB Approval of Research Project CHFS-IRB-DPH-FY07-66

Dear Mr. Gannon:


In addition to all other requirements of 45 CFR 46.101-46.409, it is the responsibility of the researcher to:

1. obtain approval by the CHFS-IRB for any modification in the research protocol or design that may increase the level of risk to a subject or a subject’s confidentiality prior to implementation;

2. advise the CHFS-IRB of any unanticipated problem involving a risk to a subject or another individual as a result of the research activity as soon as possible;

3. submit to the CHFS-IRB an electronic copy of the final research findings and conclusions; and;

4. obtain approval from the CHFS IRB prior to publication or public presentation of the findings and results of the study in order to assure human subjects protection and compliance with HIPAA regulations and confidentiality standards.
If you have any questions about any of the above, or need additional information, please contact me at (502) 564-6746 x4346 or Bob Blackburn, CHFS IRB Administrator, at (502) 564-5497 x4102.

Respectfully,

[Signature]

Doug Thoroughman, Ph.D.
Co-Chair
Cabinet for Health and Family Services
Institutional Review Board
IRB 0000 5487
FWA 0000 3302